

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2001, 17:58:39 ; Search time 1558.22 Seconds  
(without alignments)  
6352.317 Million cell updates/sec

Title: US-09-816-391A-1

Perfect score: 600  
Sequence: 1 gctggcgcgcgcgccatga.....tattcgactagtcacgcg 600

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hlg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_om:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hlgo\_hum:\*  
31: em\_hlgo\_inv:\*  
32: em\_hlgo\_rnd:\*  
33: em\_hlg\_hum:\*  
34: em\_hlg\_inv:\*  
35: em\_hlg\_rnd:\*  
36: em\_hlg\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	84	14.0	480	1	AB001381	AB001381 Streptomy
2	84	14.0	491	1	AF076983	AF076983 Caulobact
3	84	14.0	2327	1	AB033486	AB033486 Streptomy
4	84	14.0	35856	1	SCES9	AL118851 Streptomy
5	82.4	13.7	10960	1	AE005869	AE005869 Caulobact
6	75.2	12.5	20302	1	AF110185	AF110185 Burkholde
7	69.6	11.6	11691	1	AE004947	AE004947 Pseudomon
8	68.2	11.4	393	1	AF090448	AF090448 Rhizobium
9	67.4	11.2	13592	1	AE008054	AE008054 Agrobacte
10	66.8	11.1	970	1	AB032557	AB032557 Streptomy
11	66.8	11.1	42210	1	SCIC2	AL031154 Streptomy
12	66.4	11.1	11684	1	AE005903	AE005903 Caulobact
13	64.6	10.8	10851	1	AE004606	AE004606 Pseudomon
14	64.4	10.7	12116	1	AE004115	AE004115 Vibrio ch
15	63.8	10.6	3380	1	AF167159	AF167159 Sinorhizo
16	62.6	10.4	480	1	AF345628	AF345628 Pseudomon
17	61.8	10.3	568	1	ECHU2	X05994 E.coli HU2
18	61.8	10.3	11420	1	AE005631	AE005631 Escherich
19	61.8	10.3	18290	1	AE000473	AE000473 Escherich
20	61.8	10.3	176195	1	ECORH89	U00006 E. coli chr
21	61.8	10.3	318703	1	AP002567	AP002567 Escherich
22	61.6	10.3	299350	1	SME591786	AL591786 Sinorhizo
23	61.2	10.2	489	1	AF146598	AF146598 Aeromonas
24	60.2	10.0	362	1	SMG25149	U25149 Serratia ma
25	59.4	9.9	213732	1	AE001862	AE001862 Deinococc
26	59.2	9.9	2384	1	BBR132742	AJ132742 Bordetell
27	59.2	9.9	4439	1	BPE132741	AJ132741 Bordetell
28	58	9.7	145911	1	AP003014	AP003014 Mesorphiz
29	56.8	9.5	309	1	BSHBSUOM	X68448 B. subtilis
30	56.8	9.5	312	12	SYNBACHSU	M38482 Synthetic B
31	56.4	9.4	485	1	AF345629	AF345629 Pseudomon
32	55.4	9.2	1374	1	MA297483	AJ297483 Myxococcu
33	55	9.2	359	1	SM025150	U25150 Serratia ma
34	53.8	9.0	270	1	BACHDUBA	M73501 B. caldolema
35	53.8	9.0	584	1	STRHUPA	M22975 S. typhimur
36	53.8	8.9	43339	1	STYSTMEL	AF170176 Salmonell
37	53.2	8.9	10135	1	AE006017	AE006017 Caulobact
38	52	8.7	478	1	STRHUPB	X07844 Salmonella
39	50.6	8.4	270	1	BACHDUB1	M73500 B. steatrophe
40	50.6	8.4	315	1	AE0HUPA	M66574 Aeromonas p
41	50.6	8.4	429	1	BACDBPHU	D38080 Bacillus st
42	49.6	8.3	194140	1	AF242881	AF242881 Agrobacte
43	49.6	8.3	206479	1	AB016260	AB016260 Agrobacte
44	49.2	8.2	420	1	PSEHUPORA	L35257 Pseudomonas
45	49	8.2	270	1	BACHDUB3	M73502 B. caldolyt1

## ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
1	AB001381	Streptomyces lividans DNA for HSI protein, complete cds.	AB001381	1	GI:2104559	Streptomyces lividans	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Streptomyces	1 (sites)	Yokoyama, E., Doi, K. and Ogata, S.	Cloning and sequencing of the hup gene encoding the histone-like protein HSI of Streptomyces lividans
	AB001381.1	Streptomyces lividans (strain:TK24) DNA.	AB001381.1	1	GI:2104559	Streptomyces lividans	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Streptomyces	2 (bases 1 to 480)	Yokoyama, E.	Direct Submission
	AB001381	Streptomyces lividans DNA for HSI protein, complete cds.	AB001381	1	GI:2104559	Streptomyces lividans	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Streptomyces	16-JUN-1999		





```

/partial
/genes="SCE59.01c", possible oxidoreductase (partial), len:
/note="SCE59.01c, possible oxidoreductase (partial), len:
>116 aa; similar to N-terminal region of TR:Q9Z8N9
(EMBL:AL034492) Streptomyces coelicolor putative
oxidoreductase subunit SC6C5.06, 296 aa; fasta scores:
opt: 223 z-score: 268.2 E(): 1.6e-07; 37.6% identity in 93
aa overlap and to TR:P95637 (EMBL:U65440) Rhodospirillum rubrum
palustris 4-hydroxybenzoyl-CoA reductase Hbd subunit, 327
aa; fasta scores: opt: 193 z-score: 233.2 E(): 1.4e-05;
42.6% identity in 94 aa overlap"
/codon_start=1
/transl_table=11
/product="putative oxidoreductase (partial)"
/protein_id="CAF72188.1"
/db_xref="GI:6911972"
/transl_table="MGOREGAVGLTTHAPQAQAVTLPPTIDEAVALTALPAVPVPA
GGTDMAAVNSGOLRPALVIGRISLRMGVYQDGHMLDGLTHAMGRPDRAILLI
PALAAARAAGPPI"
complement(568..2205)
/genes="SCE59.02c"
complement(568..2205)
/genes="SCE59.02c"
/note="SCE59.02c, probable sugar hydrolase, len: 545 aa;
similar to TR:Q9ZH39 (EMBL:AF072374) Pseudomonas sp.
S9 beta-N-acetylglucosaminidase Ch10, 783 aa; fasta
scores: opt: 1378 z-score: 1537.3 E(): 0; 41.6% identity
in 512 aa overlap and to SW:HEX1_VIBFU (EMBL:U041417)
Vibrio furnissii beta-hexosaminidase (EC 3.2.1.52) Ex01,
611 aa; fasta scores: opt: 1023 z-score: 1142.5 E(): 0;
40.0% identity in 487 aa overlap. Contains Pfam match to
entry PF00728 glyco_hydro_20, glycosyl hydrolase family
20"
/codon_start=1
/transl_table=11
/product="putative sugar hydrolase"
/protein_id="CAF72189.1"
/db_xref="GI:6911973"
/transl_table="MTDLTVDANDVPIPRPNAVAQDADFGLGPRTLVADEAGHT
EMKRLTATGATGLPAPGEGGAVRSLIDELADEYRLDIADPGLRGAGPAGI
PAGAGIOLRLGLDPAFRARPLGRMRRLPLGIEAPRGRKMLDVRMPKMDGV
LQDLDMAHKNTLVHFLHITDQGRLEIKRRPLRTIGGSRSRKPRSRPLMEDEK
PHGGITQDIREIVATAARHITVLELPDPGSHQAIAVPELGNTDIDTALSV
NWTGWSGVDALENTLFTFVGEVLEFVEGDEEVELEGGKADMSAAQAQA
MAELGLADDELSQSTIOHFDTWLAGRRLRLLGMEILLEGGLAKAAASWRGAGSI
AARAGHDVVMGDEOVYLNPRDGGDEBPVIAVPTLEPYRREPVALPPEESG
HYTGQANLTFEVTNENRQYDLPRLAARFVMSALPAPADRDYAFERRMATHY
RLDLAAGVATRPAGRPFRPGRVGLGRPMDFPPNK"
complement(1096..1803)
/genes="SCE59.02c"
complement(1096..1803)
/note="Pfam match to entry PF00728 glyco_hydro_20,
glycosyl hydrolase family 20, score 159.70, E-value
4.4e-56"
complement(2210..3055)
/genes="SCE59.03c"
complement(2210..3055)
/genes="SCE59.03c"
/note="SCE59.03c, possible sugar transporter inner
membrane protein, len: 281 aa; similar to TR:CAF59590
(EMBL:AL132622) Streptomyces coelicolor putative sugar
transporter inner membrane protein SCP11.13, 285 aa; fasta
scores: opt: 730 z-score: 883.7 E(): 0; 42.1% identity in
278 aa overlap and to TR:AAF11008 (EMBL:AE001988)
Deinococcus radiodurans ABC transporter, permease protein,
MalP family DRI436, 283 aa; fasta scores: opt: 645
z-score: 781.7 E(): 0; 38.2% identity in 272 aa overlap.
Contains Pfam match to entry PF00528 BPD_transp.
Binding protein-dependent transport systems inner membrane
component and match to Prosite entry PS00402
Binding protein-dependent transport systems inner membrane
comp sign. Contains possible hydrophobic membrane spanning
regions"
/codon_start=1

```

	misc.feature	/transl_table=1 /product="putative sugar transporter inner membrane protein" /protein_id="CA872190.1" /db_xref="GI:6911974" translation="MNLRLARVRPRMLAEVSLALLAVVAFLYMWLVSAFPAAGEIESEPPWTLTPSIDSRFRVEEOQERPREINLIYAGSVIASIALAAVLRFRERRPTLLIMFLVMQPVKALITPLFMRFPGQLNGSILHIFAIISDFPALIMGEFYKAIVPEALEEAYIDGASTRRPKMTLPPIVLPGLYNVSEFSIRWNDFLRAKSFTLSSTOSLTPEMALVFYRKPDDPDMGVMASTWTIPVLVEFVLRRLVSGLGAVKD"		
	misc.feature	complement(2342..2563) /gene="SCB59_03c" /note="pfam track to entry PF00528 BPD transp. Binding-protein-dependent transport systems inner membrane component, score 58.90, E-value 1.1e-13"		
	misc.feature	complement(2474..2560) /gene="SCB59_03c" /note="PF00402 Binding-protein-dependent transport systems inner membrane comp sign" complement(3052..4002) /gene="SCB59_04c" complement(3052..4002) /gene="SCB59_04c" /note="SCB59_04c, possble sugar transporter inner membrane protein, len: 316 aa; similar to TR:CAB59589 (EMBL:AL132662) Streptomyces coelicolor putative sugar transporter inner membrane protein SCFI.12_302 aa; fastascores: and to TR:CAB5533 (EMBL:M117322) Streptomyces coelicolor putative binding-protein-dependent transport protein SCFI.13_317 aa; fasta scores: opt: 527 z-score: 611.6 E(): 1.2e-26; 33.1% identity in 302 aa overlap.		
	gene	Contains pfam track to entry PF00528 BPD transp. Binding-protein-dependent transport systems inner membrane component. Contains also possible hydrophobic membrane spanning regions /codon_start=-1 /transl_table=-1 /product="putative sugar transporter inner membrane protein"		
	CDS	/protein_id="CA872191.1" /db_xref="GI:6911975" translation="MSSGGVRCGRVRRNRGLRSRSPPRFGSSSTMPLYTAPALVALIGALVTVPYOGLISFILETOAOVSGSPAPAFGGNAELFGDEFNRLATVFPFAACVYSTLAVCALAVALLTRVARVRLAIMGAGMATPATVGSMVLDPDFGPVNRMIGLDHSMTGYRGSAFILLILEVVAGSPFFMVATVYGRIAPSEVLEAADGDGSQRIMRSYLAPHMRPLIVVTIOSVIMDKRVLFQIIYVMINGGISIAQNMYLNANYAQ		
Query Match	14.0%	Score 84;	DB 1;	Length 35856;
Best Local Similarity	60.4%	Pred. No. 1.2e-09;		
Matches 157;	Conservative	0;	Mismatches 100;	Indels 3;
Gaps	1;			
OY	210	tgaactggttcgaaagatgcggccagaagtccaactctgcaagtgctcgacggaggctgc	269	
Dd	8478	TGAgCTGTtGCGCCGCCGCtGCGCGCACCGCCGAGAgtTGtACCcCACAgaACGCGGaGCGCGT	8419	
OY	270	tgttaacgcccttcacaagatgtgttgtctga---ggctaigaagtcgcyggaagsgctga	326	
Dd	8418	GCTGGCGCGCTTTGGCGCGAGgTTCGCGGCaCATGTCttCCAAGGGCGAcGAGAAGTCAC	8359	
OY	327	gtcacggcctgttcttcgcgtgaagcgcttaagcgcccggctgcgacacggcgcaacc	386	
Dd	8358	CATCCCCGCGCTTCTTGACTTTCGAGCGGACACCCACCGTCCGCTCCACCGCCGCAACCC	8299	
OY	387	ggcgacctggcagcagattgatcatctcggtctcaagcgcttgatatcgcgtgctc	446	
Dd	8298	GCAAGACGGCGACGCGAATCCAGATTCCGGCGCTACAGCCTCAAGATGTCCTCGCGGCTC	8239	
OY	447	cctgcctgaagaagagcgctca	466	
Dd	8238	CAAGCTCAAGAGAAGACCGCGCA	8219	

RESULT	5
LOCUS	AE005869/c
DEFINITION	10960 bp DNA BCT 28-MAR-2001
ACCESSION	AE005869
KEYWORDS	Caulobacter crescentus section 195 of 359 of the complete genome.
ORGANISM	AE005869.1 GI:13423417
REFERENCE	Caulobacter crescentus.
REFERENCE	Caulobacter crescentus.
REFERENCE	Bacteria: Proteobacteria; alpha subdivision; Caulobacter group; Caulobacter.
REFERENCE	1 (bases 1 to 10960)
REFERENCE	Nieman, W.C., Feldblyum, T.V., Laub, M.T., Paulsen, I.T., Nelson, K.E., Eisen, J., Heidelberg, J.F., Alley, M.R.K., Ohta, N., Maddock, J.R., Potocka, I., Nelson, W.C., Newton, A., Stephens, C., Padake, N.D., Ely, B., Deboy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L., Haft, D.H., Kolonay, J.F., Smit, J., Craven, M., Kouri, H., Shetty, J., Berry, K., Uteback, T., Tran, K., Wolf, A., Vamathevan, J., Ermolaeva, M., White, O., Salzberg, S.L., Venter, J.C., Shapiro, L. and Fraser, C.M.
REFERENCE	Complete genome sequence of Caulobacter crescentus
REFERENCE	Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
REFERENCE	2 (bases 1 to 10960)
REFERENCE	Nieman, W.C., Feldblyum, T.V., Paulsen, I.T., Nelson, K.E., Eisen, J., Heidelberg, J.F., Alley, M.R.K., Ohta, N., Maddock, J.R., Potocka, I., Nelson, W.C., Newton, A., Stephens, C., Padake, N.D., Ely, B., Laub, M.T., Deboy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L., Haft, D.H., Kolonay, J.F., Smit, J., Craven, M., Kouri, H., Shetty, J., Berry, K., Uteback, T., Tran, K., Wolf, A., Vamathevan, J., Ermolaeva, M., White, O., Salzberg, S.L., Shapiro, L., Venter, J.C. and Fraser, C.M.
REFERENCE	Direct Submission
REFERENCE	Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
REFERENCE	Medical Center Dr. Rockville, MD 20850, USA
REFERENCE	Location/Qualifiers
REFERENCE	1. 10960
REFERENCE	/organism="Caulobacter crescentus"
REFERENCE	/db_xref="taxon:69394"
REFERENCE	131. 207
REFERENCE	/gene="tRNA-Asp-1"
REFERENCE	/product="tRNA-Asp"
REFERENCE	131. 207
REFERENCE	/gene="tRNA-Asp-1"
REFERENCE	272. 348
REFERENCE	/gene="tRNA-Asp-2"
REFERENCE	/product="tRNA-Asp"
REFERENCE	272. 348
REFERENCE	/gene="tRNA-Asp-2"
REFERENCE	407. 1420
REFERENCE	/gene="CC1957"
REFERENCE	407. 1420
REFERENCE	/note="identified by glimmer2; putative"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	/product="conserved hypothetical protein"
REFERENCE	/protein_id="AAK23932.1"
REFERENCE	/db_xref="GI:13423418"
REFERENCE	/translation="MRHLADMTSALPISVLDLAPVQSGSVQALRNSIDLAQHE
REFERENCE	TIGRRITWAEHNNMPGTASATAVVIOIAATTTIRVSGGVALPHAVLWAEQF
REFERENCE	GLNLALYPERIDLGRAFGTQATMRLRYAAYDSFADVVELLHWKRPASTDS
REFERENCE	VVAVEGEGDDYVITLIGSSGTGACLAALGAPFAFAHAFDALIEALTYRHRFS
REFERENCE	ESLDPRYAVVCGICACADPTDEAKRLSTGCOOFALVGRGRGLPPVDIRHASP
REFERENCE	GELAGLHTLQVAAIGSPETVRKRDYRLLELTGADLEMAAQIHDAHARRSYTILAS
REFERENCE	LRA"
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"
REFERENCE	/note="identified by match to protein family HMM"
REFERENCE	/codon_start=1
REFERENCE	/transl_table=11
REFERENCE	gene
REFERENCE	CDS
REFERENCE	complement(1417..2635)
REFERENCE	/gene="CC1958"
REFERENCE	complement(1417..2625)
REFERENCE	/gene="CC1958"

```

gene      7971..8591
CDS       /gene="CC1962"
          /note="identified by match to protein family HMW"
          /codon_start=1
          /transl_table=11
          /product="C1A protein"
          /protein_id="AAK23937.1"
          /db_xref="GI:13423423"
          /translation="MTGEMGHPLVAPDQGLTWKDSFNAFLIKWRAPRRSGYL
LTPALITVFDNRKGRKAAARQPLKGTVAQIENDAAAFKAPRPSLLRDAVAW
RWRKAGAMVIVTASPDILVAFPAAGLADLILGRLNCSDGRLGGLDGNCRK
EKVIRLEFGDVRITAVAGTSGDEMLAIADKGYRIFRGKPA"
          complement(8670..9302)
          /gene="CC1963"
          /complement(8670..9302)
          /gene="CC1963"
          /complement(8670..9302)
          /note="identified by match to TIGR protein family HMW
          TIGR00493"
          /codon_start=1
          /transl_table=11
          /product="ATP-dependent Clp protease, proteolytic subunit"
          /protein_id="AAK23938.1"
          /db_xref="GI:13423424"
          /translation="MMYDVSSTAMNLVPMVVEQTSKGERAPDIFSLKERTIFLGP
VEDGASLILAQLEFLESENPKKEIAMYNISGGVYTAGIAYDQYKSPVTCM
GMAAMGSLILAGAAGORISLPMNRIWHOPSGGPGQASDIERHAEDIIKTRRLN
EIVYHCGRTVEVERITDRDHPMSADEKAKAGLVYDSDMAENAGM"
          complement(9480..10838)
          /gene="CC1964"
          /complement(9480..10838)
          /gene="CC1964"
          /complement(9480..10838)
          /note="identified by match to TIGR protein family HMW
          TIGR00115"
          /codon_start=1
          /transl_table=11
          /product="trigger factor"
          /protein_id="AAK23939.1"
          /db_xref="GI:13423425"
          /translation="MOIYKSGEGLSRVGVTPASELARTLREARIYAEPOMYKGF
RGRPTAHVRLYKKAIMEGVIEQALNMTTKVLEDNKLRAQOPELNPSSDMKVI
AGEDLSFDLAVEMPEPIDPISLELVKPYKSDVEQELDELAKQARTYERK
EKSLAKGQDLIDVGTIDVEFAGAEAEVLGSGQPIPGEDDLGAKPEDD
VYKAKPPEEYQAKDLAGKDAEFAKVOGAEVAVDGADELAKGLGSLDALPEDD
KSNLAGRYDSSRFKRLALLDPTKHDFPLPRPVDAEFAGIMQOVASDARAGLP
PEDAKTEQDLDEYKKIAERRVRLGLVAFISGRKNDVYVTDDELDAIMREARQGA
CAQVFDKTRQADLQALRLARPIYEDKVDLIFGAKIIEKYSKDELLEEDDLPEGY
GQ"

BASE COUNT      1854 a      3619 c      3396 g      2091 t
ORIGIN
Query Match      13.7%; Score 82.4; DB 1; Length 10960;
Best Local Similarity 56.7%; Pred. No. 3.2e-09;
Matches 152; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

OY 201 caacaagctgaacctgcttcgaagatcgcccaagaatccaacctgaaccaagctcagc 260
    || || || || || || || || || || || || || || || || || || || || ||
Db 3382 CACAAAAGCCGACCTGTTACGCGCATCGCCGAGAGCGCGGATCAACAAGAACCAAGC 3323

OY 261 cgaagctgctgtaacgccttcgaagatgcttcgtagagcgtatgaagtcgcgcgaag 320
    || || || || || || || || || || || || || || || || || || || || ||
Db 3322 CAAGGACGCGCTGGAAGCTTTATCGAAGCGGTACGAGACTCCGTAATCGGCGCAAGA 3263

OY 321 cctgaagctcaacgcgcctgtcttcgcgctgagcgctcaagcgccgcgcgcgcgcgcg 380
    || || || || || || || || || || || || || || || || || || || || ||
Db 3262 CGGCGCTTGTGTGGCTTGGCACTTCAAGCGCTGACCCGCGGCTGGAACCGCGCG 3203

OY 381 caaccgcgcactgagcgacagatgcacattccgcgcttcctacgagcgctcgatctccg 440
    || || || || || || || || || || || || || || || || || || || || ||
Db 3202 TATGCCCGTACCGGTAAGACCTCCCGCGTCAAGAAGCGCCCGCGCTTCCAGGT 3143

```

```

OY 441 tggctccctgctgaagaagcgctcacc 468
    || || || || || || || || || || || || || || || || || || || || ||
Db 3142 CGCGAAGGCTGTGAACTCTCCGCTCAAC 3115

RESULT 6
AF110185
LOCUS
DEFINITION
AF110185 20302 bp DNA BCT 02-AUG-1999
Burkholderia pseudomallei strain 1026b DbHb (dbhb), general
secretory pathway protein D (gsppd), general secretory pathway
protein E (gspe), general secretory pathway protein F (gsppf), gspc
(gspc), general secretory pathway protein G (gsph), general
secretory pathway protein H (gsph), general secretory pathway
protein I (gspl), general secretory pathway protein J (gspl),
general secretory pathway protein K (gspl), general secretory
pathway protein L (gspl), general secretory pathway protein M
(gspm), and general secretory pathway protein N (gspl) genes,
complete cds; and unknown genes.
AF110185
AF110185.1 GI:4139231
Burkholderia pseudomallei.
Burkholderia pseudomallei.
Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
Burkholderia; pseudomallei group.
1 (bases 1 to 20302)
Deshazer,D., Brett,P.J., Burtnick,M.N. and Woods,D.E.
Molecular characterization of genetic loci required for secretion
of exoproducts in Burkholderia pseudomallei
J. Bacteriol. 181 (15), 4661-4664 (1999)
9350433
2 (bases 1 to 20302)
Deshazer,D., Brett,P.J., Burtnick,M.N. and Woods,D.E.
Direct Submission
Submitted (30-NOV-1998) Microbiology and Infectious Diseases,
University of Calgary, 3330 Hospital Dr. NW, Calgary, Alberta T2N
4N1, Canada
FEATURES
source
Location/Qualifiers
1..20302
/organism="Burkholderia pseudomallei"
/strain="1026b"
/db_xref="taxon:28450"
complement(1..458)
/note="orf1, similar to Escherichia coli Orf o688 and
Haemophilus influenzae gene HI1534"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD05169.1"
/db_xref="GI:4139232"
/translation="MRSARLSGRANRLRGLAENARFRRLPTPGTGLSRFAAFS
PSPFLFMTDRIVPATLVREDDGVVSPYGIYSAGALQAHNVFIRGGLPERW
RHERAFITIEFGFCNPLATWAAABRDPSCERLHRYVSERKHPAEDL"
697..975
/gene="dbhb"
/gene="dbhb"
697..975
/codon_start=1
/gene="dbhg"
/transl_table=11
/product="Dbhg"
/protein_id="AAD05170.1"
/db_xref="GI:4139233"
/translation="MNKQELDAVAAGQASRAQGTETLTLLEVIKRAVSGDSVOL
IGFGSGFGKRAARFGRNPKTETIKIPAAKVFPTAGAKAFDAVNR"
complement(1078..2421)
/note="orf8"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD05171.1"
/db_xref="GI:4139234"
/translation="MNPPLPVYVLSGLGAKTTLNHLNLRAGMKVAIVNDLAA

```



Dn	760	GCGCAAGGCTGTGGACAGCCTCTCTCGAATGATCAGAAGAGGCTGTCTCGAAAGGCAGACTCA	819
Oy	322	ctgaagctcaccgcgcttcttcgcgtctgaagcgctcaagcgccggcttcgcacggcgccg	381
Dn	820	GTTCAACTGATCTTGCTTGGGCGAGCTTCGGCTTCGGGCAAGCGCGCAAGCGCCGCCG	879
Oy	382	aaccgcgcactggcgagcacatgatgacattccgcgttccttaaggcgcttcgattccgcgt	441
Dn	880	AACCCCAAGACGGGCGAGACCATCTCAAGATCCCAGCACCAAGACGCTCAAGTTCCAGCGCT	939
Oy	442	ggctccctgctgaagaagcgctcaccgagt	473
Dn	940	GCGAAGCGGCTTCAAGACGACGACTGAACAAGCG	971
RESULT	7		
LOCUS	AE004947/c		
DEFINITION	Pseudomonas aeruginosa PAO1, section 508 of 529 of the complete genome.		
ACCESSION	AE004947		
VERSION	AE004947.1		
KEYWORDS	GI:951665		
SOURCE	Pseudomonas aeruginosa.		
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas		
REFERENCE	1 (bases 1 to 11691)		
AUTHORS	Stover,C.K., Pham,X.O., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S.L., Hutnagle,W.O., Kowalik,D.J., Lagrou,M., Garber,R.L., Goltzy,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y., Brody,L.K., Coulter,S.N., Folger,K.R., Kas,A., Lardig,K., Lim,R., Smith,K., Spencer,D., Wong,G.K., Wu,Z., Paulsen,I.T., Reizer,J., Saler,M.H., Hancock,R.E.W., Lory,S. and Olson,M.V.		
TITLE	Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen		
JOURNAL	Nature 406 (6799), 959-964 (2000)		
MEDLINE	2043737		
REFERENCE	2 (bases 1 to 11691)		
AUTHORS	Stover,C.K., Pham,X.-O.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S.L., Hutnagle,W.O., Kowalik,D.J., Lagrou,M., Garber,R.L., Goltzy,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y., Brody,L.K., Coulter,S.N., Folger,K.R., Kas,A., Lardig,K., Lim,R.M., Smith,K.A., Spencer,D.H., Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saler,M.H., Hancock,R.E.W., Lory,S. and Olson,M.V.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA		
FEATURES	Location/Qualifiers		
source	1..11691		
	/organism="Pseudomonas aeruginosa"		
	/strain="PAO1"		
	/db_xref="taxon:287"		
gene	76..1485		
	/gene="PA5346"		
CDS	76..1485		
	/gene="PA5346"		
	/codon_start=1		
	/transl_table=1		
	/product="hypothetical protein"		
	/protein_id="AAG08731.1"		
	/db_xref="GI:951666"		
	/translation="MTFAASADAPHAPOVLIOLAKITACREVSASFLPARRAH QAVLDAVGTLVLFPQSOLDARLETETGRKLIAKPERLEMSKHEHLRLPALP PTLVSPCLYDERLKEPERLIESGPILIEIASNDRGLGKASARFEEPLENIIN LMRDPPDDRAEISQAVAFTRARIQQELQTETIEPLPQAOKIKLRVPNASVDEI LTGVETEPALAAQVSWASPYAPAKISVEDAIYRVIGFDIVIMIALGALKLSK LSLPDPOQAPFYMOAQIIYAIVAEGLTRAMPRLREGSIGLGNFGYLVAKSL HVPHPSEILCRHNLEVNPHIGHAYIECHLLGITRROIGAMRWAMDMDDIYCALRF ODPSYTGNNATYANILICTTNLLRNRSIGDGTOAIPALYERLGITPERKAGDAVKKK LEAEALKEILAQENRPFG"		
	complement(1557..1949)		
gene			

gene  
CDS  
/gene="PA5347"  
complement(11557..1949)  
/gene="PA5347"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG08732.1"  
/db\_xref="GI:9951667"  
/translation="MKRRLLIMLGLMAKASRDNPAPFOKEBKDLYPOLHLLDGV  
AKHILVRLARSKRGTHPOPASFLGRDGYGFAAMTSKPOLAFMQGIQNKDIOIQ  
GNPLVYMFQCLVYLYLKKKAPKKAA"  
complement(2085..2357)  
/gene="PA5348"  
complement(2085..2357)  
/gene="PA5348"  
/codon\_start=1  
/transl\_table=11  
/product="probable DNA-binding protein"  
/protein\_id="AAG08733.1"  
/db\_xref="GI:9951668"  
/translation="MKRPELAAALAEKADLTKEQANVLLNLLDEITGALNRDSVTL  
VGFQFLQRHGAGTAKGNPQGPVKIKASNTVAFKRGKALRDVNV"  
complement(2559..3713)  
/gene="PA5349"  
complement(2559..3713)  
/gene="PA5349"  
/codon\_start=1  
/transl\_table=11  
/product="probable rubredoxin reductase"  
/protein\_id="AAG08734.1"  
/db\_xref="GI:9951669"  
/translation="MSERAPLVITIGTGLAGYNLARERKRLDGETPLMLITRDSRYS  
KPMLTSGSKNKDADGLMAEPGMAAEQNLNRIILTHRYTGLDGHQRIWIGEEVRY  
RDLYLAWAQAEPRIYVPEGDADALYPIINDLEDYARFQAAAGKRRVLLALGLCEVY  
ANDLSGCGYQIDVVAPEQVQVPLGHPAAAKAVAGLEGVRLGPVLLASLKAE  
GLEALISGEYIPDLVYVAGLRPTLELPAAGLAVNRGIVYDRSLRTSHANYAALG  
DCAEVLGNTLILYVPMACARALQALQTAGNSOYAVGMPVYVTPACPLVSPPRG  
MDGQVLVSGSTDKYLCDRPNAGVIGVLTGAAVNRKMLNKLPELMA"  
complement(3765..3932)  
/gene="PA5350"  
complement(3765..3932)  
/gene="PA5350"  
/codon\_start=1  
/transl\_table=11  
/product="rubredoxin"  
/protein\_id="AAG08735.1"  
/db\_xref="GI:9951670"  
/translation="MKKQCVVCGFYDEALGLPEEGIPAGTRMEDIPADWVCPDGV  
GRIDPEMTETIA"  
complement(4116..4283)  
/gene="PA5351"  
complement(4116..4283)  
/gene="PA5351"  
/codon\_start=1  
/transl\_table=11  
/product="rubredoxin"  
/protein\_id="AAG08736.1"  
/db\_xref="GI:9951671"  
/translation="MKKQCVVCGILTIDEAKGMPBEGLIAGTRMEDVPEDMLCPCDGV  
GLIDPEMTETIG"  
complement(4417..4818)  
/gene="PA5352"  
complement(4417..4818)  
/gene="PA5352"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAG08737.1"  
/db\_xref="GI:9951672"  
/translation="MHSKPVVIGREVFARVLGAREEOROGMAVSIIVDDGGPPLAL  
BRIDCAPIGVATIEKARSAALGRRETKRYEDMINGRSASFSLPILQATLEGGVYI  
LVDGGVAVAGVSGVRAEDARVARITGAEGL"  
complement(4823..6049)  
/gene="PA5353"  
complement(4823..6049)



/gene="glicp"  
 /note="PA5353"  
 complement(4823. .6049)  
 /gene="glicp"  
 /codon\_start=1  
 /transl\_table=11  
 /product="glycolate oxidase subunit glicp"  
 /protein\_id="AAG08738.1"  
 /db\_xref="GI:9951673"  
 /translation="MOTNLCDSARORAEAAERILRSYHCGFCNMTCPYORLGD  
 LDGSRIRYLKIOFLERDEASERTRLDCTSCNCTGSPGVETXKHLIDIRDL  
 AERLPGEFRORLIDGRLIAPRGILGALAGAGRIPLIPEMOKSLRVAAPR  
 RPPRRARRRILLEGCVQPALSPNTAAARVDRLISYTPAAAGCCGALDHDA  
 QERKQARRNIDAMWPLNREGAIVOTASGCGAFYKDYAYLLRDPKADKAREVA  
 AARDIVLEVALDESLIGRLCDRLAFPCITQAHQRGAGVAEVALRILGPHLTV  
 ADHLICGSAGTYSLTFQALPSQRLRDLRDALESGADVYVTNAVCGAHLIDAGRP  
 VRHIELLDALPGGE"  
 complement(6059. .7138)  
 /gene="glicp"  
 /note="PA5354"  
 complement(6059. .7138)  
 /gene="glicp"  
 /codon\_start=1  
 /transl\_table=11  
 /product="glycolate oxidase subunit glicp"  
 /protein\_id="AAG08739.1"  
 /db\_xref="GI:9951674"  
 /translation="MSOPYFDPDAGADHAGALLERVNDAIARAPRIYVGDSEKAFIG  
 RPYAGPDLAEHRGIVAPCELVITARAGTLDLQAVLARGOMLACBAGFEGR  
 ATYGCYLAGISGPRRRMACAVRDFYGLTSLIGLTHLRFGEVMKNVAGYDLRLL  
 AGSFGLGVTEVSFVLPMPKQAGISLQLPYAMLRKLAEGGQPLPSAACGHE  
 ALNRLGEGGVSARSARDLGEALDGEFVADIRQLPFFDDPRPLMRSLPALAE  
 TPLPGRLDMGALRMLKSDAPADIRAWTAHGHATCYRAGHLDAAPQPLSPGLL  
 GLHRAKRLRDPGIFNPGEMVAEL"  
 complement(7138. .8637)  
 /gene="glicd"  
 /note="PA5355"  
 complement(7138. .8637)  
 /gene="glicd"  
 /codon\_start=1  
 /transl\_table=11  
 /product="glycolate oxidase subunit glicd"  
 /protein\_id="AAG08740.1"  
 /db\_xref="GI:9951675"  
 /translation="NIIYDERLDGPLQVQDKDLAEIRLRPLDELHAAEDLRPY  
 ECDGSAYRCTLPLVALPERIEVOGILALCHRLKYPVYARGGTISGALPLENGY  
 LTVMAFRRLIETDIPGRAPRVPGVNRNALSQAAPHGLIYAPDPSSQIACSTIGNV  
 AENAGYRCLTKGTIVHNLQYDITLBEGRSLSSALDSAGFLLALFTGSEGLAG  
 VVEVTVLLPRPAAKVLILASFDDVESAGRAVADLIGAGIVPAGLEMDNLSIRAE  
 DEIHAGYPVDAAILLCGLDGYEADVEDCERVELFEAGATSVROADEAROREM  
 AGRKNAPVAGRISPDYICMGDSIPRELPRVGHGIAEVSREKGLNAVVFHAGDGM  
 HLLLEDANLAGEIEREALGRIILELCAVAVGSLTGEHSCYGERKINQCAQFNDEL  
 TLFHAKAFDPDAGLINPKKNVPTLHRCAEFGAMVHGRLLPPELERP"  
 8842. .9597  
 /gene="glicp"  
 /note="PA5356"  
 8842. .9597  
 /gene="glicp"  
 /codon\_start=1  
 /transl\_table=11  
 /product="transcriptional regulator glicp"  
 /protein\_id="AAG08741.1"  
 /db\_xref="GI:9951676"  
 /translation="MOKORVADOVAERIERLIVDGVTKGQALPSRRRLVYALGCSRS  
 ALRGLRALKRGITIDTEGRSSFYADIDRNADVSLHMLFGSPRTLLDLEVRLL  
 EGEARLALRGTEADFTLLARTEEMLSHETPTIDREHARRDHARHAISESSH  
 NPVLVHTLOSLLLELSTVFASVNNIYHRPPKRRQIDROHARLYALREROPDOARA  
 AAEHLSIRDNREITEOEORLYRATRLINGWG"  
 9677. .10213  
 /gene="PA5357"  
 9677. .10213  
 Query Match 11.6%; Score 69.6; DB 1; Length 11691;

Best Local Similarity 52.8%; Pred. No. 3.1e-06;  
 Matches 150; Conservative 0; Mismatches 134; Indels 0; Gaps 0;  
 QY 205 aagtcgacctcgcttcggaagatcgccgaagtcgaacctgcaagcgtagcgag 264  
 DB 2351 AACCCGACACTAGCCGCGCGATCGCCGAAAAGGCCGATCCACCAAGACAGCCCAAT 2292  
 OY 265 gctgctgaagccttcggaagatgctgtagagtagtaagtcgagcgagcgctg 324  
 DB 2231 CGCGTTCACACGCCCTGTTGATGAATAATCACCGCGCGTGAACCGCAAGACAGCGTG 2232  
 QY 325 aagtcacggcgctgttcctcgctgagcgcgcaagcgcccgctcgacggcgcaac 384  
 DB 2231 ACCCTGGTGGTTCGGGACCTTCTCGCAAGCCATCGCGAGCCGCGGAGAGAC 2172  
 QY 385 ccgagcactgagcgagcagattgacatccgcttcctgaagcggtatctcgctgagc 444  
 DB 2171 CCGGAGACCGGCGACCGCGTGAAGATCAAGCGCAGCAACCGTTCGCTCAAGCGGCGC 2112  
 OY 445 tccctgctgaagcgagcgctgacagcgagtagcctctgtag 488  
 DB 2111 AAGCCCTGCGCGAGCGGTCACTGACCGCCGAGACCGCGAG 2068  
 RESULT 8  
 AF090448 393 bp DNA BCT 04-OCT-1998  
 LOCUS AF090448  
 DEFINITION Rhizobium leguminosarum HU-1like protein gene, complete cds.  
 ACCESSION AF090448  
 VERSION AF090448.1 GI:3694945  
 KEYWORDS  
 SOURCE Rhizobium leguminosarum.  
 ORGANISM Rhizobium leguminosarum.  
 Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 Rhizobiaceae; Rhizobium.  
 REFERENCE 1 (bases 1 to 393)  
 AUTHORS Liu,S.T., Chang,W.Z., Cao,H.M., Hu,H.L., Chen,Z.H., Ni,F.D.,  
 Lu,H.F. and Hong,G.F.  
 TITLE A HU-1like protein binds to specific sites within nod promoters of  
 Rhizobium leguminosarum  
 JOURNAL J. Biol. Chem. 273 (32), 20568-20574 (1998)  
 MEDLINE 98352101  
 REFERENCE 2 (bases 1 to 393)  
 AUTHORS Hu,H.L. and Hong,G.F.  
 TITLE Direct Submision  
 JOURNAL Submitted (05-SEP-1998) Shanghai Institute of Biochemistry, Chinese  
 Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, China  
 FEATURES  
 source 1. .393  
 /organism="Rhizobium leguminosarum"  
 /strain="8401"  
 /db\_xref="taxon:384"  
 1. .393  
 /gene="hurle"  
 52. .327  
 /note="hurle"  
 /gene="hurle"  
 /note="hrl"  
 /codon\_start=1  
 /transl\_table=11  
 /product="HU-1like protein"  
 /protein\_id="AAC62499.1"  
 /db\_xref="GI:3694946"  
 /translation="MNKMLSAVAKRGLKRSDAASAVADVVAELNKKGDRL  
 AGFSFTVSHRAATGRNPSTGAEDVIDIARRNPKTPGKLDAVNG"  
 BASE COUNT 91 a 101 c 115 g 86 t  
 ORIGIN  
 Query Match 11.4%; Score 68.2; DB 1; Length 393;  
 Best Local Similarity 53.1%; Pred. No. 9.3e-06;  
 Matches 145; Conservative 0; Mismatches 128; Indels 0; Gaps 0;  
 OY 202 aacaaagtcgactcgcttcggaagatcgccgaagtcgaacctgcaagcgtagcgcc 261

Db	55	AACCAAGATGAACHTCGTCCGGCAGCTACCAGAAAAGGACAGGACTGCACCAAGTCGTACGCCG	114
QY	262	gaagctgcttaacgcctcccaagtgtgtcgtcgagcatgaatcgcgcgaagc	321
Db	115	GCTTCCGCTGTGAACGCGGTWTTCAGCTGTGCCAGGCTGTAACATCAAGAACAAAGGCGCAC	174
QY	322	ctgaagctcacccgccttcttcgcgtctgaagcgcgtcaagcggcgcctcgaccgcgcgc	381
Db	175	ATCCGCGCTCGGGGGTTTGCGAGCTTCACCGCTTCCTCATCTGTGCGCGCAACGAGGCGCGT	234
QY	382	aaccgcgcactgagcagcagattacatccgcgccttcctcagcgcgttcgatctcgcgt	441
Db	235	AATCCGTGACGGCGCGCTGAAGTGCACATTCCGGCTTCGCCAAGCTGCCCAATATTCACGCT	294
QY	442	ggtcccctgctaagaaggccgtcacccagatga	474
Db	295	GGCAAGGCGCTGAAGGACGCGCTCAACGCGTGA	327
RESULT	9		
AE008054			
LOCUS	13592 bp	DNA	BCT 14-AUG-2001
DEFINITION	Agrobacterium tumefaciens strain C58 circular chromosome, section 112 of 254 of the complete sequence.		
ACCESSION	AE008054 AE007869		
VERSION	AE008054.1	GI:15156309	
KEYWORDS	Agrobacterium tumefaciens. Agrobacterium tumefaciens Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.		
SOURCE ORGANISM	1 (bases 1 to 13592) Hinkle,G., Slater,S.C. and Goodner,B. Complete Genome Sequence of Agrobacterium tumefaciens C58 (Rhizobium radiobacter C58), the Causative Agent of Crown Gall Disease in Plants		
REFERENCE AUTHORS TITLE	Unpublished 2 (bases 1 to 13592) Hinkle,G., Slater,S.C. and Goodner,B. Direct Submission Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street, Cambridge, MA 02139, USA		
JOURNAL REFERENCE	Location/Qualifiers		
TITLE	1. .13592		
JOURNAL	/organism="Agrobacterium tumefaciens"		
FEATURES	/strain="C58"		
source	/db_xref="taxon:358"		
gene	238..2655		
CDS	/gene="AGR_C_2329"		
	238..2655		
	/note="ATP-DEPENDENT PROTEASE LA"		
	/codon_start=1		
	/transl_table=11		
	/product="AGR_C_2329p"		
	/protein_id="AAK87056.1"		
	/db_xref="gi:15156310"		
	/translation="MTNITSASGCTPVPLIRDIIVPPHVIPLVFGREKSITALEE VMGSDKOIMLVQTINASDDDPAPAPIHKVGTVANVLILKIPDTGVLYLVEGKAROI IDSYGRDFYEASAATPLQEPADPEVELEALSRSVSEESYLANKISPEVGAAGAO IDDYSKLADIYASHLSIKITEKOMLETFVSQDLAEALGMBEIEIIVLQVEKRISRBI VKROMEKTOREYVYNEDOMKAILOKEIGDEDRDMAELERIAKTSLKEAKEDVAE MKTRKMSPMSEAATVVRYNTDWLLGGLPWGKKSKIKNDINAETIILDODHGDLKVE RIVELYAVQARATIRGPILCLVVPPEGVGRKSLAKSTAKARGRYVMAAGVVDPAEA IRGRRRRTIGSMPCRIVOSMKAkakaknFLIDETDKMGDFRODPSALLEVEDPBO LPKAIKHEHALYEEFSDDALMAISOQYIEADVNRFERELMLARKAYEIIRKGS TSVTAVTANIIDIYGIPIREFRGAEAREDOVGVGLAWTEVGEELLTLEGVMPGKGGR MTVGNLKVEWKESGISAAASYVRRAVDGTEPRRPKSDIHVHPGAPPKGPSAG VMAATAIVSTITGILPVSKDVAMTGETILRGVPLEDIGCKEKLIALNGKITKIILPEE NAKDLAETPDNVKNEMETIPVSRNGEVIKHALIRQPEITWDGSIETPVLAIVEGVDD GNQIIAH"		

gene 2945. .3220  
/gene="AGR\_C\_2331"  
2945. .3220  
/gene="AGR\_C\_2331"  
/note="DNA-BINDING PROTEIN HRL18"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_C\_2331p"  
/protein\_id="AAK87057.1"  
/db\_xref="GI:15156311"  
/translation="MNKNELYSVAEAKGALRKADAASAVDAVFETVQGEELKNGDILRL  
AGFSGFSYRRRAASKGRNRPSTGAEDIPARNVPKFSAGKGLDAVNS"  
3234. .3398  
/gene="AGR\_C\_2333"  
3234. .3398  
/gene="AGR\_C\_2333"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_C\_2333p"  
/protein\_id="AAK87058.1"  
/db\_xref="GI:15156312"  
/translation="MRCPCAGMRLEPOAHEETCFDALMPLSGVRNPSKALSGGPVSPILM  
VIESLEAPV"  
3537. .5762  
/gene="AGR\_C\_2334"  
3537. .5762  
/gene="AGR\_C\_2334"  
/note="(AL117669) putative secreted protein"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_C\_2334p"  
/protein\_id="AAK87059.1"  
/db\_xref="GI:15156313"  
/translation="MFLDPRGIKMAFRFESHVALTAALLTSNAPFAAEVFNRIASEPP  
VAKNIPADODASVTSAEIYVASEDEGKTLIYSDPLIGITGIDITDAPKAPAGACALLIL  
OGEPIVSVAAGKAGLAVGNTSTENRKNRPGSKITVYDIARKTEASCDLGGDDSVASVR  
DTEPVAIAIENRDEEVNKGALPQMPASITLYTVGILKSDAADCATLKKIDYGLSEISRP  
EDPEEFVIAINANGELIATWQENNHIVIIDKRTGVKTHFSAGTVNLEGIDTKTDGDL  
KTTDQOEGKREPDVAKWMLDNDRIIVIANEGMOMGSGRGTIFEDTKILYEAGSLREHR  
AAALHGGPEKRSKAKGIEPEGLAATPDGQKFFVILERASVVAAYKDTGAEEQLTQ  
LAPSGISGPAVAIPIPSRNILVTANPADLAEGGASRHYMLERABGOAAYPOLVASVTV  
DPAIIGMADSGIYSDAKPKGLIYAVSPSYGNQPSITYTIDATKKPAITSVLEPKDD  
GIAAORLIDLEGITLDGKGFVILASBGDSAKIYPAHLIYNVNAKGEIKAEIALPKRLNG  
DTRFEFEGTAVMGKDDATLWMAVOREMGTDKEGKVLKLYSNPKSKEMGAVRPLDPTQ  
EAGWGVLSITTAOGDNVYIVERDNOIGDKARKIKLYRVAIADLPGKIGGLPYAVAKO  
EAKHDELPLDKAQNTGVYVDKLEGFTEPKDAGTAAYAVTNDGDVSSGETLIFTYMLKALH  
N"

gene complement(5755. .5874)  
/gene="AGR\_C\_2336"  
complement(5755. .5874)  
/gene="AGR\_C\_2336"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_C\_2336p"  
/protein\_id="AAK87060.1"  
/db\_xref="GI:15156314"  
/translation="MPHRASCKISGTAFHSGLSGDTKKKPGYARLRFSPVPLGITS"  
3615. .6105  
/gene="AGR\_C\_2337"  
complement(5968. .6105)  
/gene="AGR\_C\_2337"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_C\_2337p"  
/protein\_id="AAK87061.1"  
/db\_xref="GI:15156315"  
/translation="MDEHMSIDAFTVTLAILLPYIIFRWVQDSEKEDPFSKSSSPG  
R"  
3662. .6862  
complement(6407. .6862)  
/gene="AGR\_C\_2338"

gene 2945. .3220  
/gene="AGR\_C\_2331"  
2945. .3220  
/gene="AGR\_C\_2331"  
/note="DNA-BINDING PROTEIN HRL18"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_C\_2331p"  
/protein\_id="AAK87057.1"  
/db\_xref="GI:15156311"  
/translation="MNKNELYSVAEAKGALRKADAASAVDAVFETVQGEELKNGDILRL  
AGFSGFSYRRRAASKGRNRPSTGAEDIPARNVPKFSAGKGLDAVNS"  
3234. .3398  
/gene="AGR\_C\_2333"  
3234. .3398  
/gene="AGR\_C\_2333"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_C\_2333p"  
/protein\_id="AAK87058.1"  
/db\_xref="GI:15156312"  
/translation="MRCPCAGMRLEPOAHEETCFDALMPLSGVRNPSKALSGGPVSPILM  
VIESLEAPV"  
3537. .5762  
/gene="AGR\_C\_2334"  
3537. .5762  
/gene="AGR\_C\_2334"  
/note="(AL117669) putative secreted protein"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_C\_2334p"  
/protein\_id="AAK87059.1"  
/db\_xref="GI:15156313"  
/translation="MFLDPRGIKMAFRFESHVALTAALLTSNAPFAAEVFNRIASEPP  
VAKNIPADODASVTSAEIYVASEDEGKTLIYSDPLIGITGIDITDAPKAPAGACALLIL  
OGEPIVSVAAGKAGLAVGNTSTENRKNRPGSKITVYDIARKTEASCDLGGDDSVASVR  
DTEPVAIAIENRDEEVNKGALPQMPASITLYTVGILKSDAADCATLKKIDYGLSEISRP  
EDPEEFVIAINANGELIATWQENNHIVIIDKRTGVKTHFSAGTVNLEGIDTKTDGDL  
KTTDQOEGKREPDVAKWMLDNDRIIVIANEGMOMGSGRGTIFEDTKILYEAGSLREHR  
AAALHGGPEKRSKAKGIEPEGLAATPDGQKFFVILERASVVAAYKDTGAEEQLTQ  
LAPSGISGPAVAIPIPSRNILVTANPADLAEGGASRHYMLERABGOAAYPOLVASVTV  
DPAIIGMADSGIYSDAKPKGLIYAVSPSYGNQPSITYTIDATKKPAITSVLEPKDD  
GIAAORLIDLEGITLDGKGFVILASBGDSAKIYPAHLIYNVNAKGEIKAEIALPKRLNG  
DTRFEFEGTAVMGKDDATLWMAVOREMGTDKEGKVLKLYSNPKSKEMGAVRPLDPTQ  
EAGWGVLSITTAOGDNVYIVERDNOIGDKARKIKLYRVAIADLPGKIGGLPYAVAKO  
EAKHDELPLDKAQNTGVYVDKLEGFTEPKDAGTAAYAVTNDGDVSSGETLIFTYMLKALH  
N"

complement (6407. .6862)  
/gene="AGR\_C.2338"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_C.2338p"  
/protein\_id="AAK87062.1"  
/db\_xref="GI:15156316"  
/translation="MSSRFHDKRISFLLIGNNRTFKHDCCKAPFALHNENKLV  
HSSRPQAQVYKAVRASPPGPAKRPVAVGFLPLAFRGKSCPEKQNFKALAVF  
RSAYITATDAADMIKRTTRRAISSVGNAPRLHRCGREFESLIAMHFLKTS"  
7577. .7942  
/gene="AGR\_C.2340"  
7577. .7942  
/gene="AGR\_C.2340"  
/note="NADH DEHYDROGENASE I CHAIN A 1 (NADH-UBIQUINONE  
OXIDOREDUCTASE CHAIN A 1)"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_C.2340p"  
/protein\_id="AAK87063.1"  
/db\_xref="GI:15156317"  
/translation="MTELLSYVYIPIAFIPIGALVIGLALLITAPFAVAAYKADPDKLSN  
PFCGPAAPDAAMKPDIREYLVSLFIIPDEVALFLPWAYSVFGEMGFWGSMVFL  
AVLTIGFTIEKKKALEMA"  
7912. .8517  
/gene="AGR\_C.2341"  
7912. .8517  
/gene="AGR\_C.2341"  
/note="NADH DEHYDROGENASE I CHAIN B 1 (NADH-UBIQUINONE  
OXIDOREDUCTASE CHAIN B 1)"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_C.2341p"  
/protein\_id="AAK87064.1"  
/db\_xref="GI:15156318"  
/translation="MEESGACGAGMGNSTLVAPQPKIIDPATGKPYGSDAYFTEI  
NDELAKGEFLVSTDELITWARTGSLMMQFGLACCAVEGLMASGPRYIDMERGVAP  
RASPSQSDVMTAVGTILTNKMAPALKRYDQMEPRYVISMGSANGGYYHYSYAAVVR  
GCDRVVPIYVPGGPTPAEALLYGVLLILQKRIKRTGTIER"  
8539. .9141  
/gene="AGR\_C.2342"  
8539. .9141  
/gene="AGR\_C.2342"  
/note="(A)245398) nuoc1"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_C.2342p"  
/protein\_id="AAK87065.1"  
/db\_xref="GI:15156319"  
/translation="MSEALNDLDAVYKEARGLSVASDIAYGELTNTPEENYALLT  
FLRDVQGVGFAMIIDICGVDPQREKREDVYHLISPRONLRVGIKLOVADEGVSS  
TPVYQAFMEFEAMIDMGPIPEGKDLRLITDYGREGHPLKRDPPVYGVFEVRIDY  
VLRVLVYPVVLKQEFNPFSLSPREGTEYVLPDEKAKQ"  
9138. .9566  
/gene="AGR\_C.2344"  
9138. .9566  
/gene="AGR\_C.2344"  
/note="hypothetical protein SPAC29B12.13 - fission yeast  
(Schizosaccharomyces pombe)"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_C.2344p"  
/protein\_id="AAK87066.1"  
/db\_xref="GI:15156320"  
/translation="MSAPRAREHQICLCGAVGFATIGAREFVYCHCSMCRRNSGAF  
MAVECTIDAEENORLGVYSSKMGECFCNCGSTSGKSGKHFAVSLQAFKPS  
SFTVYQQLTDEKSSYSFAQTONTMGPEFIAMITSAEH"  
9568. .10758  
/gene="AGR\_C.2346"  
9568. .10758  
/gene="AGR\_C.2346"  
/note="NADH DEHYDROGENASE I CHAIN D 1 (NADH-UBIQUINONE

[illegible]

	Query Match	11.18;	Score 66.8;	DB 1;	Length 970;	
	Best Local Similarity	57.08;	Pred. No. 1.8e-05;			
	Matches 122;	Conservative	0;	Mismatches 92;	Indels 0;	Gaps 0;

  

ORIGIN	214 a	319 c	331 g	106 t
OY	256 cagcgcgagctgtctgttaacgcttccagagatgttccttgtagcgcatgaatccggc	315		
Db	326 CAGCGCCGCGACGGCTGTGCAGCGCGGCTCCTTGACGCCCTTCGTCGGTCCGCGGAC	385		
OY	316 gaaggccgaagaactcaacgcgctgtctccgcgtgagcgctcaagcgccgcgctcgacc	375		
Db	386 GACC GGCTCTGGTCGTCACCGGCTTCGCTTCCTTCGGAAGAAGTGCAGCCGCCGCGCTTAC	445		
OY	376 gagccgaacccgcgcactgycgagcagattgacattccgcttctaagcgcttcgatc	435		
Db	446 GCCCGCAACCCGCAAGACGGCGAGGGGTTCCGGGTCAGAAAGACCTCCGTCGCCGCTTC	505		
OY	436 tcgcctgctccctcgtcgtcgaagaagcgcttacgg	469		
Db	506 CGCGCGGGCCAGGGCCTTCAGAGACTCTGTGGAGCG	539		

  

RESULT 11	SCIC2	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS
	SCIC2	42210 bp	DNA	BCT	15-JAN-1999	
	Streptomyces coelicolor cosmid 1C2.					
	AL031124.1	GI:3355667				
	3-isopropylmalate dehydratase large subunit; 3-isopropylmalate dehydrogenase small subunit; 3-isopropylmalate dehydrogenase; branched-chain amino acid aminotransferase; carbonyl transferase; delta-1-pyrroline-5-carboxylate dehydrogenase; glx; glutamyl-tRNA synthetase; histone-like DNA binding protein; hydrolase; llyE; leuB; leucD; leuD; lyase; secreted lyase; transfer-RNA-Gln; transfer-RNA-Glu; ureAB; urease alpha subunit; urease beta and gamma subunits; ureC.					
SOURCE ORGANISM	Streptomyces coelicolor A3(2).					
REFERENCE AUTHORS JOURNAL REFERENCE AUTHORS TITLE	JOURNAL					
Direct Submission	Submitted (23-JUL-1998)	Streptomyces coelicolor sequencing project Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK				
3 (bases 1 to 42210)						
Redenbach,M., Kleser,H.M., Denapate,D., Eichner,A., Cullum,J., Kinash,R. and Hopwood,D.A.						
A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome						
Mol. Microbiol. 21 (1), 77-96 (1996)						
97000351						

Notes:

Streptomyces coelicolor sequencing at the Sanger Centre is funded by the BBSRC.

Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.

(URL: <http://www.sanger.ac.uk/projects/S.coelicolor/>) CDS are numbered using the following system eg SC7B1.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

```

FEATURES
SOURCE
1. .42210
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:1002226"
/clone="cosmid 1c2"
1. .1228
/genes="SC1C2.01"
<1. .1228
/genes="SC1C2.01"
/noise="SC1C2.01", probable delta-1-pyrroline-5-carboxylate
dehydrogenase, partial CDS, len >408 aa; similar to many
e.g. TR:O50443 (EMBL:AL010186)
delta-1-pyrroline-5-carboxylate dehydrogenase (M.
tuberculosis) (543 aa), fasta scores: opt: 1753 z-score:
2214.2 E(): 0, 65.1% identity in 407 aa overlap, and
PURT_HUMAN delta-1-pyrroline-5-carboxylate dehydrogenase
(563 aa), fasta scores: opt: 1273 z-score: 1730.4 E(): 0,
48.5% identity in 408 aa overlap. Contains PS00687 and
PS00070 Aldhyde dehydrogenases glutamic acid active site
and cysteine active site and Pfam match to entry PF00171
aldhch. Aldhyde dehydrogenases, score 114.70, E-value
1.8e-30"
/codon_start=2
/transl_table=11
/product="delta-1-pyrroline-5-carboxylate dehydrogenase"
/protein_id="CA19968.1"
/db_xref="GI:3355668"
/db_xref="SPTRMBL:086502"
/translation="IDSPCELIDPFRVHYARNILAEOPANSOGVYNNRMDHRLPLEG
EVAITPFRNSAIANLPTPALMGVYVYKRPSTQTHRAVLIMQLEALPLPGYIN
LVTGDGSADEVALEHRDLGIFHTGSTKTFQHLMKTVGNIEKRTIPRVGEGGK
DLFVADHPADSAVALEKTLATRGAEFYQOQCSATSRAYIPASIMNDGEFEERAEVDYI
TMGVDLTLSNFIIGAVIDERSPAKKAAIDRAKEDETCIIVAGSGVDSGVFVRVTV
ECQPEHVEFTEFEGFLAVHYVDDSDADYADMLTQMESVSDYALTGVSISNDRAAP
AAYMKELRYAAGNFYINDKSTGAVVGGQPRGGRASGTNDKAGAPQIMRWTLTRA
KETLVAPFDITTPHMG"
1. .3807
/noise="true overlap with cosmid 8D9"
2. .1189
/genes="SC1C2.01"
/noise="Pfam match to entry PF00171 aldheh. Aldhyde
dehydrogenases, score 114.70, E-value 1.8e-30"
464. .487
/genes="SC1C2.01"
/noise="PS00687 Aldhyde dehydrogenases glutamic acid
active site"
548. .583
/genes="SC1C2.01"
/noise="PS00070 Aldhyde dehydrogenases cysteine active
site"
1246. .1283
/noise="hairpin_loop with 18bp stem"
stem_loop

```

```

RBS      1522..1527
         /note="possible RBS upstream of SCIC2.02"
gene      1537..1668
         /gene="SCIC2.02"
CDS       1537..1668
         /gene="SCIC2.02"
         /note="SCIC2.02, questionable ORF, len: 43 aa"
         /codon_start=1
         /transl_table=11
         /product="hypothetical protein SCIC2.02"
         /protein_id="CAI19970.1"
         /db_xref="GI:3355669"
         /db_xref="SPTREMBL:O86503"
         /translation="MAETTVRRRVRLHSTSDSKRNAALQRLDRDNGATGH
1744..1787
         /gene="leuB"
         /gene="leuB"
         /transl_table=11
         /codon_start=1
         /product="3-isopropylmalate dehydrogenase"
         /protein_id="CAI19970.1"
         /db_xref="GI:3355670"
         /db_xref="SPTREMBL:O86504"
         /translation="WRSRLNLAIVIPGDSIGGEVVAEGLKIVLSAVLPQDVKLETKEPDF
GARRHAAGETLTLDALNALKADAILGAIIGDPSVPGVLERGFLKLIRAFPHHVA
LRPSKLIGVATPIAGOEIDFVVVRETEEGTITNGTIRKGEHEVATSVNTAVN
GVERVVRAPAAQAPRKKLTIYHKNNVLFPAGLITWIRNKVAEAPETTPHYLV
DAATFETVDPARFPIVTDNLFQDITDLAAVAGGICVAAISNINSGDFPMPFV
VHGSRPDIAGCKADPPIATVIVSALLIHLGITEDEARIEDAVSHDLSERGDLPKRS
SEIGTILARVAG"
misc_feature 1759..1766
         /gene="leuB"
         /note="Pfam match to entry PF00180 isodh, Isocitrate and
isopropylmalate dehydrogenases, score 356.60, E-value
1.3e-105"
misc_feature 2458..2517
         /gene="leuB"
         /note="PS00470 Isocitrate and Isopropylmalate
dehydrogenases signature"
RBS       3038..3042
         /note="possible RBS upstream of llvE"
gene      3052..4140
         /gene="llvE"
         /gene="llvE"
         /transl_table=11
         /codon_start=1
         /product="branched-chain amino acid aminotransferase"
         /protein_id="CAI19971.1"
         /db_xref="GI:3355671"
         /db_xref="SPTREMBL:O86505"
         /translation="WTFPTILKPSAHPLSDSERAAILANEGFGRHFTDHWITKWE
GRGMDGLVYATPSIDPATMVLYHKOETIEGLKAVRRPDGSVATPRPKNGARFQA
SSRRHMGELVVDTEIEACDALVAODEKVVYAHGSESLYLRPMIATVEGLRPA
EYLFVYISPAAGFPGCVKPVSLIWSEDRYRAVPGMGAKTGGNTAAISLAQAA
AKGDGVYDAIRKVEEIGKMTYFYVGNKTYTPSLNGSILEGTRDLSLTAVD
LGYVAEBRVSVDQMQRDSENGTILTEVFACITAAVITPVGVKRAAGQWQSGSETGE

```

```

misc_feature
3763..3852
/locus="VTORLADLIDIQFCTVADPHGMHTLA"
/gene="tlive"
/note="PS00770 Aminotransferases class-IV signature"
gene
complement(4214..5644)
/gene="SC1C2.05C"
complement(4214..5644)
CDS
/gene="SC1C2.05C"

Query Match 11.1%; Score 66.8; DB 1; Length 42210;
Best Local Similarity 57.0%; Pred. No. 1.2e-05;
Matches 122; Conservative 0; Mismatches 92; Indels 0; Gaps 0

OY 236 cagcgccgagcgcgcgcgttaacgcgcctccagagatglttcgtcagagctagaatcgcgc 315
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41235 CAGCGCGCGGAGCGGTGTGACAGCGCGGTGTGACAGCGCGGTGTGCGTCCGTCGCGCGGCG 41294
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY 316 gaagcctgaagctaccgcgcgcctgtcttcgcgtgagcgcgttaagcgcgcgcgcgcgcac 375
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41295 GACCGGGCTCGTCCGTACCGCGCTTCGGTTCCTTCGAGAAAGTGCACCGCGCGCGCGCTTAC 41354
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY 376 ggcgcacacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 435
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41355 GCCCGCAACCCGCAACACGGCGCGCGCGGTTCGGGTCAAGACACCTCCGTCCCGCGCTTC 41414
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY 436 tcgcgtgcctccctgcgtgaagaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 469
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41415 CGCGCGGCGCGAGGCGCTTCAGAGACCTCGTGACGC 41448
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
AE005903 11684 bp DNA BCT 28-MAR-2001
LOCUS
DEFINITION Caulobacter crescentus section 229 of 359 of the complete genome.
ACCESSION AE005903
VERSION AE005903.1
KEYWORDS GI:13423852
SOURCE
ORGANISM Caulobacter crescentus.
Caulobacter crescentus.
Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
Caulobacter.
1 (bases 1 to 11684)
Niernan,M.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E.,
Eisen,J., Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R.,
Potocka,I., Nelson,M.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B.,
Deboy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
Halt,D.H., Kolony,J.F., Smit,J., Craven,M., Khouri,H., Shetty,J.,
BERRY,K., Uterback,T., Tran,K., Wolf,A., Vamathevan,J.,
Emolaevev,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and
Fraser,C.M.
Complete genome sequence of Caulobacter crescentus
Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
21173698
2 (bases 1 to 11684)
Niernan,M.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J.,
Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I.,
Nelson,M.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B.,
Laub,M.T., Deboy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
Halt,D.H., Kolony,J.F., Smit,J., Craven,M., Khouri,H., Shetty,J.,
BERRY,K., Uterback,T., Tran,K., Wolf,A., Vamathevan,J.,
Emolaevev,M., White,O., Salzberg,S.L., Shapiro,L., Venter,J.C. and
Fraser,C.M.
Direct Submission
Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers
1..11684
/organism="Caulobacter crescentus"
/db_xref="taxon:69394"
complement(73..1236)
/gene="CC2328"
complement(73..1236)
/gene="CC2328"

FEATURES
source
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS

```

/note="identified by match to PFAM protein family HMM  
PF00441"  
/gene  
/codon\_start-1  
/transl\_table-11  
/product="acyl-CoA dehydrogenase family protein"  
/protein\_id="AAK24299.1"  
/db\_xref="GI:13423853"  
/translation="MSGFWLLEQEAIREGVANVCAFDDEYWRRTDGEFPEAFVA  
AIAEGWLVGAMPEVNGAGLGLTEAAVMQSVAGSGFSGASAIHLNIFGMPILVK  
FGDEORNLRLPRIISGEDKCFAPVPPNSGLDTSLETRAEKEGGYRLNGRIWYT  
GADRAKNIILITRTTPEKOCAPKPGHLSFTYDRENTAEAKPIPMGRKAYECMNLFE  
DLVPPADLVGEEGKQYILHGINPERVIFPVAETGCGAALAKATTYANERYER  
PIQNGVAPHLAKSMWELPANLAFKASALTDQGEKCGAANAANTYLDAGEGFHAC  
EAVLVAGHMGVYKEVDYERFPEAMTARIAPVSRREMIINFIARVIGLPKSY"  
/gene  
/complement(1265..2281)  
/gene="CC2329"  
/complement(1265..2281)  
/gene="CC2329"  
/note="identified by match to PFAM protein family HMM  
PF01266"  
/gene  
/codon\_start-1  
/transl\_table-11  
/product="D-aminic acid oxidase family protein"  
/protein\_id="AAK24300.1"  
/db\_xref="GI:13423854"  
/translation="MKLRGARIAVAGAGALGSNAVLRLAQRGFRTVDPSPGDANA  
SSVAGMLAVSEALPDPASRNHLDIKRRADLPDPARYLGLIEDRGVRIEISDWM  
RSVAVSLGALVEPAGAFIAEDWIDRAGLALRGASAGAVFRLAVCGFAGVL  
ELADGGEPPDTLIATPGILOGEITPPTMRGILRLAAGCDADAPVVRGSG  
VYLAPGVAVAGATMEGRDVLPLEATYGLIRRAALRDLDLESATYEVGVRYT  
PDGLPLGVMSRTPEGMVLAVGARRNGWLFAPLVDMVAAYLFGDNLGPDALAMARRKE  
KPYT"  
/gene  
/complement(2495..2902)  
/gene="CC2330"  
/complement(2495..2902)  
/gene="CC2330"  
/note="identified by match to protein family HMM"  
/codon\_start-1  
/transl\_table-11  
/product="transcriptional regulator, Cro/Ci family"  
/protein\_id="AAK24301.1"  
/db\_xref="GI:13423855"  
/translation="MNDPRAEQAHFVLYVGARLRIRRRKMGISQTVADALGTTQ  
QIKYERGANRISKLYDAKALQAPVSTFBELEDPDGVDGFAQRATEFVSTPE  
GLELASLPRLADRLRRRVLDVRAVDDAP"  
/gene  
/complement(3209..3487)  
/gene="CC2331"  
/complement(3209..3487)  
/gene="CC2331"  
/note="identified by match to protein family HMM"  
/codon\_start-1  
/transl\_table-11  
/product="DNA-binding protein HU"  
/protein\_id="AAK24302.1"  
/db\_xref="GI:13423856"  
/translation="MNVSDLYDAVADDEKLTKQAALIIDGVFKSISDAVAGEEVS  
IPGGKFKVQAKPRTGRNPTATITIAASKKGFPAKQKDAVNG"  
/gene  
/complement(3609..4838)  
/gene="CC2332"  
/complement(3609..4838)  
/gene="CC2332"  
/note="identified by Glimmer2; putative"  
/codon\_start-1  
/transl\_table-11  
/product="conserved hypothetical protein"  
/protein\_id="AAK24303.1"  
/db\_xref="GI:13423857"  
/translation="MLDMAYDLDRKLAISDAKYDASCASSGAVORNSUNGKYS  
TEGGICHTAIPDRCTISLKLITLNFCTIDCAICINRISNTFRAPFTVOEVDLTL  
NFKRNYIEGFLTSSGVIRSGDYMEEMVRAKSLRDHDFRGYIHLKLIPEASPELA  
AEGGLADRVASINTELPRDSDIKALAEKDPDTDRAMGRRLAIDHAEGRARR  
AFASOSTOILVGADEAKDGIILSRANLVCAYRLSRVYSAFSPIDPASRLDLP  
PLIREHLTYQADWIMRFYGFASQSEIIVAAGDLDLSVPKTSMAHRDRPVDQIAD

gene  
/complement(4838..6277)  
/gene="CC2333"  
/complement(4838..6277)  
/gene="CC2333"  
/note="identified by match to TIGR protein family HMM  
TIGR00758"  
/gene  
/codon\_start-1  
/transl\_table-11  
/product="phage SpoI DNA polymerase-related protein"  
/protein\_id="AAK24304.1"  
/db\_xref="GI:13423858"  
/translation="MQVRLASEVDPAQWRRAARVIRACGAPESLWTVTERELFND  
EPYTDATTFVYPAAFEMAOQVYILNRSIDRFLLRLIMRLREPRLEINPADQMA  
RADMAKAVSRAHKKKAFFVRVEDDAKETAYARVTEATAPFAARFSNM  
NMTILNDVCVANGERLWYSEGDADPADSEDAQALMTYVASTFNPRLRPMR  
OEMPKRYVRLPEALILPGLIEAADRAAMATTPRPSERYLKAORARAPYNA  
GAPTTIDAVRAGVSVCRCDLREATQVPGDGAPSAPLMFVGEQGDDEDLGSLY  
GPAQILFDRLAEVSVRREQTVYTNVVKHFKHELNGKRRLRHKTPTQGEVSACRMWIDA  
ERRLVPRVTVIMLGATAGVGMRPPTVRETGRPLPLPDGVOGLVTFHPSYLLRLPD  
AAARENAYSFVDLRLAGRLAGLAATSP"  
/gene  
/complement(6418..6494)  
/gene="trna-his-1"  
/product="trna-his"  
/complement(6418..6494)  
/gene="trna-his-1"  
/complement(6418..6494)  
/gene="CC2334"  
/complement(6418..6494)  
/gene="CC2334"  
/note="identified by Glimmer2; putative"  
/codon\_start-1  
/transl\_table-11  
/product="hypothetical protein"  
/protein\_id="AAK24305.1"  
/db\_xref="GI:13423859"  
/translation="MNGKRPVETPPGSLTFVAYTQALLRDLERDGHATYDGRAVE  
GLIHPEQAQADSRILAEAHVRRFGHRAAFQSALSRHGRARVATADHPISVARI  
VPGHVFTRGLTGVTGRDRTGSRVRRNRGRDRHRSRLFPLENGDRORRRH  
VRRSHRRARLNGRRRRRRTIGRRRRRRRRRRRRRRLVYVLLHEHFEILL  
ROISEOORVOTSHDGGDTKFTLLATVAFRRERIVAAMLGHTLLSHASPPGYHNY  
RRAPRYRSRYPGCSPP"  
/gene  
/complement(7265..7799)  
/gene="CC2335"  
/complement(7265..7799)  
/gene="CC2335"  
/note="similar to SP:005606 GB:X70139 PID:49194;  
identified by sequence similarity; putative"  
/codon\_start-1  
/transl\_table-11  
/product="ExbD/TolR family protein"  
/protein\_id="AAK24306.1"  
/db\_xref="GI:13423860"  
/translation="MAKISGGGDRFNVNONSEINVTPEVDMLVLLITFYAAPLA  
SVSEVETLPPAVAKSSPPSPKPYVYSIKOSGLYGDNEISIDELGIDIKKNGRRDP  
SKKRITRADEKRYKGFMEVYNTLDDNGYTSVALYGRDSDQ"  
/gene  
/complement(7805..8713)  
/gene="CC2336"  
/complement(7805..8713)  
/gene="CC2336"  
/note="identified by match to protein family HMM"  
/codon\_start-1  
/transl\_table-11  
/product="MocA/TolQ/ExbD proton channel family protein"  
/protein\_id="AAK24307.1"  
/db\_xref="GI:13423861"  
/translation="MLDIRKTPFPIALVGMALMASAPAFADAAAAPATAEAOPAT  
DAAAAPADAAAAPAEAPAPAEAPAAHSLTFPICMEQACAVYVYVYVIGVLSV  
ESVLLTGLTFEPAALNKOSDKFLAEARGAKSLADRISSSEGBNPAVDAALAS  
QVEVLISQALVAAGEHRDLSRAYAIINAVOSLAKRLSSGQVFLASVSGSGPTG  
LFTVVGIMTSFGLIANTNTNLAVVAFGAEALATATGIGLFAIPAVIVIFYNYFQTRI  
SAGTNRSGEVAELMNAISRLDKGA"

[illegible][illegible]

gene  
/codon\_start=1  
/transl\_table=1  
/product="ATP-dependent Clp protease ATP-binding subunit  
ClpX"  
/protein\_id="AAG05191.1"  
/db\_xref="GI:9947785"  
/translation="MTDRNGEDNGKLLCYSCFGKQHEVRKLLAGSPVFCIDCCVDL  
CNDILREVEVOQAQSSGCHKLPAPKEITFLIDQYVIGGERKATLVAAVHYHKLKNO  
BKDKDIDELGKSNILMIGTGPSTKLLAETLAPLALVETADTAPLAVVEQVEDE  
NITQCLLQKCDYDKEAKMGITYYIDEIDKISKSNPSBITDQSGEYQALMLLEG  
TVASVPQGGKRNHPQOEFLLQYDTRNILEITCGAGFGLKRLSSSARQISFNAVRES  
QDMGKRVGAEEVEPEDLVKFGLLPEEVGRPLPVATLDELIDENALMQILTEPNALTL  
KQYALFMEGEGDLEFRPDALKAVARKLERTGKARGSRISILEGILLDTWYEIPSOOD  
VSKYVIDESVIDGSSQPLMIYENSEKPAKAAPEA"  
5871..8267  
/gene="lon"  
/note="PA1803"  
5871..8267  
/gene="lon"  
/codon\_start=1  
/transl\_table=1  
/product="Lon protease"  
/protein\_id="AAG05192.1"  
/db\_xref="GI:9947786"  
/translition="KKTIVELPLPRVYVYPRHWIPLTYGKESITALEAANTGGR  
QTLIAQNPADDDEGDELKMGTVATVLDLKLPGDTAVVLEGEORGVERFIEE  
EHIRHAAQALIDANVAGERAEVFTSRLLSQEOYVQAGKVPAAEVSLSSIDERS  
LYDVAHAHLKLEIQKODILEITDLSVEHYLLALDAEIDLQVEKTRIRGVKQHE  
RSORRYLVNEOKKALIQELGDIDEGHNEVEELKRIDAAGLTKEAHTAPALKEK  
MSPAAEAATVNSYIDMLLVNPMWAKESVRHDLAKAEIDLADHGYLEEYERLLEVI  
AVOKRKLKAGVCLVGLVGPVCGKSLAESIARBNKRPVMAIGVADPEAEITGHRP  
TYIGMSPRLQKMTKTVYRNPFLIDETDKGSDMRGDPASALLEVIDPEQNINPD  
HLEVDYDLSDDVLCITANSMTIPALDLDRMEVILKPEYTEDEKAVNTAKLLTRQVQ  
ANGILKGELEFEELRDILIRYTRYEAGVRSLEQRIYAVCRKAAVEHAHKLRIQAVS  
STELNRYLVGRKFRYGLAEQODQIGOVYGLMTQVGGELITEAVAVGKGLTRQES  
LGDVAATELITAVVRSRQSSIGAIAPFHEKTRILHVPAGATPKDPSGIMCTGA  
LYSATQIPIPRADVAMTGEITLRGOVLAIGGLKEKLLAHRGGIKTVIIPENYRDLK  
ELHPNIKSDLVIKPKYVIDEVLQALQYAPETLPDAPEWAMDEKREPDSEKHSITH  
"

gene  
8403..8675  
/gene="hnpB"  
/note="PA1804"  
8403..8675  
/gene="hnpB"  
/codon\_start=1  
/transl\_table=1  
/product="DNA-binding protein HU"  
/protein\_id="AAG05193.1"  
/db\_xref="GI:9947787"  
/translition="MNKSELDIAAASADIPKAVAGRALDAVIESVTGALKAGDSVIL  
VQGFPAVKEBAARFGRNPQGTGKPIKIAAAMPKPKAKALKDAVN"  
8727..8803  
/product="rRNA-asp"  
8907..10772  
/gene="ppid"  
/note="PA1805"  
8907..10772  
/gene="ppid"  
/codon\_start=1  
/transl\_table=1  
/product="peptidyl-prolyl cis-trans isomerase D"  
/protein\_id="AAG05194.1"  
/db\_xref="GI:9947788"  
/translition="MLOINRDSOGMIKTTIGVITVLSLTGDAIIRATHDSNVAA  
KNGDIDISNEVOQAVDMGRQLLDQRKDDPPSLMDKLKEAKGLIERTDILLDA  
AADDFAPSDALDQLILQTPFEOYDGFNADRFQYIRQNNYSMVFQMLQGEMLIL  
QGLRGLAGGEGVTNDELQSFARLEKQTRDPATLAKDAKSSVSDDVEVAFEGHKK  
SEFTPEQVAVVEYVELKSSSFEDQYKVOEQLQALQSEIANTLSEQRDAHILLEVND  
KQGDQAAKIDELIARLAKGEDFALAKERSODIGSAATGSDGYAGRGVYDAFAFEE  
ALYAKOGVSAVPKTPYGYHLIKLLGVADPEVPSLTIKRLLEDLQKQWBERFVEE  
ATKDLSSAYEADLSQPAQEMGLKVQTSOPRGSRGCGIATANQIYQAPASAVILEE  
ANSGAIGTAPGVVAVLKRCHNRKKEQPLEQVQANITRERLAEKAAAEAKRGREALIL  
AELNGRRTSSAAGTESKAVEAASRGHECVDRLLQAVFRMQRPAPAKQPPSSVGYTLA

BASE COUNT	2225 a	3228 c	3439 g	1959 t	
ORIGIN	GDYVIRLNGVSEEEAASIDSEKAMRYRFLASRSGQADFAFRQLQAKAEVEXY"				
Query Match	10.8%	Score 64.6;	DB 1;	Length 10851;	
Best Local Similarity	52.0%;	Pred. No. 4.6e-05;			
Matches 145;	Conservative 0;	Mismatches 134;	Indels 0;	Gaps 0;	
OY	202	aacaagcttgacctcgtttcgaagatcgcccaagaagtcacacttaccagaagctcaagcc	261		
Db	8406	AACAAGTGGGAAGTCAATGATGATGCAATTCACCGATTCGTGATGATTCGCCGAACCTGTGGC	8465		
OY	262	gagctgtctgtaaacgccttcacagatgtgtctgcgcgaggtcgtatgaagtcgcgcgaagc	321		
Db	8466	GGTCCGCACTGAGACGAGTGAATGCAATGCAATGCGTCAATGGGCGCCCTTAAGGCTGTACTCTCC	8525		
OY	322	ctgaagctcacccgcctgtttctcgcgtgaagcgcgtcaagcgcgcgcgttcgcacgcgcgc	381		
Db	8526	GTCGTCGTGGTGTGGTTCGGACCTTCCTTCGTCAAAAGAGCGCGCTGCTCGCACCGGTCCG	8585		
OY	382	aaccgcgcactgagcgcagatgaacatccgcgttcctacgggttcgtatctccgct	441		
Db	8586	AACCCGCAACTGTGCAACCCGATCAAGATGCCGCGTCCGCAAGATCCCGGCTTCAAGCC	8645		
OY	442	ggtccctcgtcgaagaagcgcgtcacgcgaagtacctct	480		
Db	8646	GGTAAAGCCCTGAAGATGCTGTCAATCAAGCGCTTTT	8684		
RESULT 14					
AE004115	LOCUS 31-JUN-2000				
DEFINITION	Vibrio cholerae chromosome I, section 23 of 251 of the complete chromosome.				
ACCESSION	AE004115 AE003852				
VERSION	AE004115.1 GI:9654674				
KEYWORDS	Vibrio cholerae.				
SOURCE	Vibrio cholerae.				
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.				
REFERENCE	1 (bases 1 to 12116)				
AUTHORS	Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwin,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L., Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D., Ermolaeva,M.D., Vamathevan,J., Bess,S., Qin,H., Dragoi,I., Sellers,P., McDonald,L., Utterback,T., Fleischmann,R.D., Nierman,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R., Mekalanos,J.J., Venter,J.C. and Fraser,C.M.				
TITLE	DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae				
JOURNAL	Nature 406 (6795), 477-483 (2000)				
MEDLINE	2 (bases 1 to 12116)				
REFERENCE	Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwin,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.A., Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D., Ermolaeva,M.D., Vamathevan,J., Bess,S., Qin,H., Dragoi,I., Sellers,P., McDonald,L., Utterback,T., Fleischmann,R.D., Nierman,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R., Mekalanos,J.J., Venter,J.C. and Fraser,C.M.				
AUTHORS	Direct Submission				
TITLE	Submitted (14-JUN-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA				
JOURNAL	Location/Qualifiers				
FEATURES	1..12116				
SOURCE	/organism="Vibrio cholerae"				
	/strain="N1961"				
	/serotype="O1"				
	/db_xref="taxon:666"				
	/chromosome="I"				
	/note="Biotype: El Tor"				
	complement(167..1400)				
	/gene="VC0265"				
gene					





```

/note="identified by Glimmer2; putative"
/codon_start=1
/translation_table=11
/product="hypothetical protein"
/protein_id="AAF3449.1"
/db_xref="GI:9654684"
/translation="MTTPRLNINSAFLKWKQSLFLLSLILAGCSSTPNPNEIQE
NQPLFGKRTIGDTSFYWTTSFQKPSASLYTSDGYGYSQSYNMEGQREIRREB
EHLGKGLKIVYPRHAIKRSKGEENVYQYIRAGGVLPDMNEQLAHVLDKAAVETK
ODKGLKLIQYVYNGKFLGKGVLPDMNEQLAHVLDKAAVETK"
/geneID=LLDDDAGHCVCPEPBL"
/geneID=LLDDDAGHCVCPEPBL"
/gene="VC0275"

```

Query Match	10.7%;	Score 64.4;	DB 1;	Length 12116;
Best Local Similarity	51.8%;	Pred. No. 5,1e-05;		
Matches 146;	Conservative 0;	Mismatches 136;	Indels 0;	Gaps 0;

QY 202 aacaagtcgacccctcgttctcgaagatcgcccaagaatgccacctgaccaaagcttcagcc 261  
||||| - - - ||||| - - - ||||| - - - |||||  
Db 7900 AACAGACCCCAATTGATCGACTTATCGCAGAGAAAGCGACCTGACATAAAGTGCACGCG 7955

**Oy**    262 gaggctgctgtttaacgccttcacagatgtgttctgcgaagctataagaiccggcgcaagc 321  
       | | | | | | | | | | | | | | | | | | | | | |  
**Db**    7960 AAAGCAGCTCTTGAAGCGACTCTGGGTGCGGTTGAAGGTGCTCTGAAAGACGGCGACCAA 8019

Oy 322 ctgaagctcaacgagcctgttctccgctgagcgcgctcaagcgccgcgctcgacaaccggccg 381  
| | | | |  
Db 8020 GTACAACGATTTGGTTTTGGTACATTCAAAGTTAATCATCTCGTTCTGCCTGTACTGGCCGT 8079

Oy 382 aaacgcgcaactggcgagcgaattgacatccgcgtcttctacggcggttcgtatctccgt 441  
||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 8080 AATCCAAGACCGGTGAAGATCAAAATTGGGCGCAGCTAAGCTTCCATGCATTTCGTAGCA 8139

Qy	442	ggtctccctgcttaagaagcgctgacacggatgaccttgc	483
Db	8140	GGTAAGCGCTGAAGAAGCGCATCAAGTATTCGTTTACT	8181

RESULT	15
AF167159	
LOCUS	
DEFINITION	
	AF167159 3380 bp DNA BCT 20-APR-2000
	<i>Smorhizobium meliloti</i> lon protease (lon) and histone-like protein (hupB) genes, complete cds.

SOURCE	ORGANISM
Slinorhizobium meliloti.	Slinorhizobium meliloti
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group.	Rhizobiaceae, Slinorhizobium.

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 3380)	Summers, M. L., Botero, L. M., Busse, S. C. and McDermott, T. R.	The $\alpha$ -inositol:phosphatidylinositol 4- $\alpha$ -phosphatase is involved in regulating exopolysaccharide synthesis and is required for nodulation of

J. Bacteriol. 182 (9), 2551-2558 (2000)  
20225855  
2 (bases 1 to 3380)  
Summers, M.L.  
Direct Submission  
Submitted (08-JUL-1999) Land Resources and Environmental Sciences  
Montana State University, 334 Leon Johnson Hall, Bozeman, MT  
59717-3120, USA

```

FEATURES
source
location/qualities
1..3380
/organism="Sinorhizobium meliloti"
/db_xref="taxon:382"
350..2770
/gene="lon"
350..2770
/gene="lon"
/codon_start=1

```

BASE COUNT  
ORIGIN

758 a 971 c 1037 g 614 t

gene

CDS

transl\_table=11  
/product="ion protease"  
/protein\_id="AA05300.1"  
/db\_xref="GI:6175841"  
/translation="MTNKISPTAESATYEVLPGLDIVPEPHMIVPLFVGRKSTRALAE  
EWGTDQIMLVYQINATDDPEPSATYKGTIANVQLKLPDGTAKVVEGRSRAE  
IEETRPDDPEVEAMAHALPEPDEPVEIETLSSRVSEFYSYFLANKIIPREVGVAS  
QIDYDSKLADTVASHLSIKIYEKOEMETLTSYKMRLEKALGFMEGELISYLOVEKRIS  
RIVROMEKQREYLIINOMKAIQIOKEIGSDSDGEMLEBERISKTSEARERKADA  
ELKIKRSPMSMAAPVYRVYLDMLGILPMGKSKSIITDILNHAKEVLTDPHFGIDYK  
ERIVERIYAVARBSKIKGPIPLCYGPPGVQTSIAKIANATGEYIRMALEGVDRDA  
EIRGHRRTYIGSPMGKVVQSMKKRKNPPLFLDEIDKQDGFQDSSALLEVLDPE  
QNSTPEMDHLYEVEDLSNMEITTTANTLNPPLMDMEVFRLAGYEEERLEETLAKR  
LEKRALRDHALQTNESVTDGALTALIONTPREGVSNPRELRLAKRAVETELKRS  
TKREYSAENINIDYLVGPRRHEAEEDSDGVVYTGAMTEVEGELLITIGVAMPGKG  
RMVGTGLRDVWKESISAAASYVRSIADGIEPLDEKRDIDHIVHEGATPKNGPSGA  
GVMAATIVISVMTGIPISKDVAMTGETTLGRLVPIGGLEKILLALRGGIIKRYLIP  
ENAGSTINDIPDWNKNSFELLIPVSRMGVIAHALRLPEIEMDPASPALPVSQSD  
EAKGSIH"

2992. 3264  
/gene="hupB"  
2992. 3264  
/gene="hupB"  
/note="HupB"  
/cdon\_start=1  
/transl\_table=11  
/product="histone-like protein"  
/protein\_id="AA05301.1"  
/db\_xref="GI:6175842"  
/translation="MNKNELYAAVADKAGLSKADASSAVAFETIOGELKNGGDIYRL  
VGGNGFVSRRASKGNPSTGAEVDIIPANNVPKFTAGKGLKQAVN"

Query Match	10.6%	Score 63.8	DB 1	Length 3380
Best Local Similarity	52.4%	Pred. No. 8e-05		
Matches 140	Conservative	0	Mismatches 127	Indels 0
				Gaps 0

QY 202 aacaagctgcacctcgtttcgaagatcgcaccagaftccaaactgaccaaagtctcagcc 261  
||||| ||| ||||| | ||| ||| ||| ||| ||||| ||  
Db 2995 AACAAAATGAGCGCTGTGGTCCGCCGTAAGCCGACAGGCGCGACTCTGAAGGCAGATGCC 305

QY 262 gagcgctgcgttaacgacctccagagtgtgttcgtcgaagcatgaatccggcggaagc 321  
||||| |||| | | | | | | | | | |  
Db 305 TCTTCTGCCGTGATGCAGTTTTTTGGACCACTTCAGGGCGGAAGAAGCGGGCGCAG 311

Qy 322 ctgaagctacccgcctgttctcgcgtgagcgcgtaagcggcgcgtgcaccggccgc 381  
+ ||| ||| + ||| ||| ||| ||| |||  
Db 3115 atccgccctcgttgcccttcggcaattttctccgatcggccgcgcaggctgacaaaggccgc 317

OY 382 aaaccgcgcaactcgtcgagacgaattgcatctccggcttcctaaggcgttctgatctccgct 441  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 3175 AACCGTCACGGGATGCCGAATTGCATTCGGCACGCCAACGTGCCGAACTTACGGCC 323

QY 442 ggtctcctgtgaagaagccgtacc 468  
||| | ||| | ||||| |  
Db 3235 GGCAAGGCGCTCAAGAGCGCGTCAAC 3261

Search completed: December 20, 2001, 19:47:34  
Job time: 6535 sec

Search completed: December 20, 2001, 19:47:34  
Job time: 6535 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2001, 18:54:44 ; Search time 167.05 seconds  
(Without alignments)  
3079.288 Million cell updates/sec

Title: US-09-816-391a-1

Perfect score: 600  
Sequence: 1 gcctggcgccgagccatga.....tattccgactagcttcagcg 600

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N.Geneseq.1101.\*  
1: /SID8/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID8/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID8/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID8/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID8/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID8/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID8/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID8/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID8/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID8/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID8/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID8/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID8/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID8/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID8/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID8/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID8/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID8/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID8/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID8/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID8/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID8/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	47.2	7.9	349980	21	AAAF21609
C 2	47.2	7.9	349980	21	AAAF21610
C 3	47.2	7.9	1437668	21	AAA81490
C 4	46.8	7.8	8093	21	AAA81512
C 5	45	7.5	62909	22	AAAF28545
C 6	43.6	7.3	16075	20	AAV99811
C 7	43.2	7.2	16079	8	AAAF70753
C 8	40.6	6.8	273	21	AAAF79661
C 9	39.6	6.6	1026	21	AAA88001
C 10	39.6	6.6	1575	15	AAAF0417
C 11	39.6	6.6	1578	15	AAAF0418

12	39.6	6.6	1638	17	AAAF9320
13	39.6	6.6	1638	22	AAAF27164
C 14	39	6.5	1326	22	AAAF4356
C 15	39	6.5	1326	22	AAAF48225
C 16	39	6.5	1326	22	AAAF48225
17	37.8	6.3	276	22	AAAF90714
18	37.8	6.3	276	22	AAAF90714
19	37.6	6.3	858	20	AAZ11393
20	37.4	6.2	969	20	AAAF98059
C 21	37.4	6.2	1017	20	AAAF98059
22	37.4	6.2	42235	20	AAAF98059
23	37	6.2	2233	20	AAAF16480
24	36.6	6.1	77536	21	AAAF14651
25	36.2	6.0	300	22	AAAF90842
26	36.2	6.0	375	20	AAAF98061
27	35.8	6.0	2025	20	AAAF98060
28	35.4	5.9	1822	11	AAAF06095
29	35.2	5.9	683	21	AAAF13147
30	35.2	5.9	47981	22	AAAF30757
31	35	5.8	1950	19	AAAF23479
32	35	5.8	1950	21	AAAF13897
33	35	5.8	3246	18	AAAF93610
C 34	35	5.8	17612	19	AAAF23494
C 35	35	5.8	17612	21	AAAF13905
C 36	35	5.8	17612	22	AAAF30870
C 37	34.8	5.8	2721	12	AAAF10212
C 38	34.8	5.8	2721	12	AAAF10543
C 39	34.8	5.8	8438	15	AAAF73500
40	34.6	5.8	624	17	AAAF06312
C 41	34.6	5.8	2001	22	AAAF60958
C 42	34.6	5.8	10732	21	AAAF10594
C 43	34.4	5.7	1524	12	AAAF10499
44	34.4	5.7	1525	21	AAAF91207
45	34.2	5.7	471	20	AAAF15020

#### ALIGNMENTS

RESULT 1	AAAF21609/C	AAAF21609 standard; DNA; 349980 BP.
ID	AAAF21609	
AC	AAAF21609	
XX		
DT	13-MAR-2001	(first entry)
XX		
DE	Neisseria meningitidis B nucleotide sequence SEQ ID NO:110.	
XX		
KW	Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine; diagnosis; antigen; detection; infection; gene therapy; antibacterial; ds.	
XX		
OS	Neisseria meningitidis.	
XX		
PN	WO20006791-A1.	
XX		
PD	09-NOV-2000.	
XX		
PF	08-MAR-2000; 2000WO-US05928.	
XX		
PR	30-APR-1999; 99US-0132068.	
PR	08-OCT-1999; 99WO-US23573.	
PR	28-FEB-2000; 2000GB-0004695.	
XX		
PA	(CHIR) CHIRON CORP.	
PA	(GENO) INST GENOMIC RES.	
PI	Piazza M, Hickey E, Peterson J, Tettelin H, Venter JC, Maignant V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V, Rappelli R, Frazer CM, Grandi G;	
PI	WPI; 2000-647603/62.	
XX		
DR		

Cholesterol oxidase  
Streptomyces preg  
Human GAT4 nucle  
Heart muscle cell  
Human GAT4 codin  
CPE 14 coding sequ  
CPE 13 coding sequ  
M. vaccae antigen  
Nucleotide sequenc  
Nucleotide sequenc  
Nucleotide sequenc  
SAB virus gB glyco  
Nucleotide sequenc  
2CPE 33 coding seq  
Nucleotide sequenc  
Nucleotide sequenc  
Sequence encoding  
Aspergillus oryzae  
Microsporidia meg  
Pseudomonas xpcp s  
Pseudomonas alcali  
Mycobacterium tube  
Pseudomonas xpcp s  
Pseudomonas alcali  
BamHI J-I fragment  
BamHI J-I fragment  
DNA encoding Pseud  
Thaumatin II synth  
P. putida KT2440-a  
Gene encoding a su  
Beta-ketothiolase  
Beta-ketothiolase  
Tritichoderm reesel



RESULT 3  
 AAA81490/c  
 ID AAA81490 standard; DNA; 1437668 BP.  
 AC AAA81490;  
 XX  
 XX  
 DT 04-DEC-2000 (first entry)  
 DE  
 XX N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.  
 XX  
 XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KM antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KM Meningococcus B; MenB; ds.  
 XX  
 XX Neisseria meningitidis.  
 OS  
 XX WO20002430-A2.  
 PN  
 XX 20-APR-2000.  
 PD  
 XX 08-OCT-1999; 99WO-US23573.  
 PF  
 XX 09-OCT-1998; 98US-0103794.  
 PR 30-APR-1999; 99US-0132068.  
 PR  
 XX (CHIR ) CHIRON CORP.  
 XX  
 XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
 PI Mesigian V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
 PI Rappuoli R, Pizza M;  
 PI  
 DR WPI: 2000-318079/27.  
 XX  
 XX Isolated nucleotide sequences of Neisseria meningitidis which can be  
 PT used in the diagnosis and treatment of N. meningitidis infection and  
 PT other Neisserial infections, for example, N.gonorrhoea -  
 PS  
 PS Claim 7; Page 866-1272; 1760pp; English.  
 XX  
 XX The present invention describes methods of obtaining immunogenic  
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
 CC represent specifically claimed Neisseria meningitidis genomic DNA  
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to  
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to  
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a  
 CC medicament) for treating, preventing or diagnosing infection due to  
 CC Neisserial bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against Meningococcus B; against all serotypes;  
 CC and/or against all pathogenic Neisseriae. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.  
 CC  
 CC Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;

Query Match 7.9%; Score 47.2; DB 21; Length 1437668;  
 Best local Similarity 48.8%; Pred. No. 0.0097;  
 Matches 127; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

OY 202 aacaagctgacctgcttgcgaagatgcgccagaagttccaaactgaccgaagctcgaagcc 261  
 |||||  
 Db 1205593 AACAAAGCTTGAAATGATGCAAGACGATTGCTCAAGAACCCGACATTTCAAGCGCCGCA 1205534  
 OY 262 gaggctgctgtaaacgctccaccagatgtgttcgtcgaagctatgaagctcgcgaagc 321  
 |||||  
 Db 1205533 CAANAAGCTTTGGATGCGCATACCAATGCAATGCAACACCCCTCGAACAAGCGGACAC 1205474  
 OY 322 ctgaagctcaccgagctgttctccgctgaagcgctcaagcgccgagctcgaagcgccg 381  
 |||||  
 Db 1205473 GTTACCTTGGTTCGGTTCCGCTACTTCTTACGTGGCGGCAAGTCGGAACCAAGCGCCG 1205414  
 OY 382 aaccgagcactggcgagcagattgacattccgcttcctcagcgcttcgagctcgtatcgct 441  
 |||||  
 Db 1205413 AACCCCAAAACCGGCGAGCTCTGACCATTTGCCGCCCAAAAAGCCTTAATTCGCGCC 1205354  
 OY 442 ggtccctcgtcgaagaaagc 461  
 |||||  
 Db 1205353 GGCAAGCTTTGAAGACGC 1205334  
 RESULT 4  
 AAA81512  
 ID AAA81512 standard; DNA; 8093 BP.  
 XX  
 XX AAA81512;  
 AC  
 XX  
 DT 04-DEC-2000 (first entry)  
 DE  
 XX N. meningitidis partial DNA sequence gnm\_59 SEQ ID NO:59.  
 XX  
 XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KM antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KM Meningococcus B; MenB; ds.  
 XX  
 XX Neisseria meningitidis.  
 OS  
 XX WO20002430-A2.  
 PN  
 XX 20-APR-2000.  
 PD  
 XX 08-OCT-1999; 99WO-US23573.  
 PF  
 XX 09-OCT-1998; 98US-0103794.  
 PR 30-APR-1999; 99US-0132068.  
 PR  
 XX (CHIR ) CHIRON CORP.  
 XX  
 XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
 PI Mesigian V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
 PI Rappuoli R, Pizza M;  
 PI  
 DR WPI: 2000-318079/27.  
 XX  
 XX Isolated nucleotide sequences of Neisseria meningitidis which can be  
 PT used in the diagnosis and treatment of N. meningitidis infection and  
 PT other Neisserial infections, for example, N.gonorrhoea -  
 PS  
 PS Claim 7; Page 1372-1375; 1760pp; English.  
 XX  
 XX The present invention describes methods of obtaining immunogenic  
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
 CC represent specifically claimed Neisseria meningitidis genomic DNA  
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to  
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to  
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a

medicament) for treating, preventing or diagnosing infection due to  
 CC Neisserial bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against Meningococcus B; against all serotypes;  
 CC and/or against all pathogenic Neisseriae. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.

SO Sequence 8093 BP; 1993 A; 2045 C; 2032 G; 2022 T; 1 other;

Query Match 7.8%; Score 46.8; DB 21; Length 8093;  
 Best Local Similarity 48.5%; Pred. No. 0.0013;  
 Matches 126; Conservative 1; Mismatches 133; Indels 0; Gaps 0;

OY 202 aacaagctgactcgttctgaagatcgcccaagaagctcaacctgaccagagctcagcc 261  
 |||||  
 DB 77 aacaagctgactcgttctgaagatcgcccaagaagctcaacctgaccagagctcagcc 136  
 |||||  
 OY 262 gagcgtcgttgaagcgttcacagatggtgtcgtcgaagctagaagctcgcggaagc 321  
 |||||  
 DB 137 caaagaagcttgatgac 196  
 |||||  
 OY 322 ctgaagctcagcgcgtctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 381  
 |||||  
 DB 197 gtaacttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt 256  
 |||||  
 OY 382 aaccgcgcacgtcgc 441  
 |||||  
 DB 257 aaccgcgcacgtcgc 316  
 |||||  
 OY 442 ggcctccgtcgtcgaagaagc 461  
 |||||  
 DB 317 ggcgaagcttgaaagagc 336  
 |||||

RESULT 5

AAF28545/c  
 ID AAF28545 standard; DNA; 62909 BP.

AC AAF28545;

DT 04-APR-2001 (first entry)

DE Genomic fragment #32.

XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;

KW bronchopulmonary; endocarditis; meningitis; ss.

OS Moraxella catarrhalis.

XX WO200078968-A2.

PN 28-DEC-2000.

PF 16-JUN-2000; 2000MO-US16649.

XX 18-JUN-1999; 9905-0140121.

PA (INCY-) INCYTE GENOMICS INC.

PI lagace RE, Patterson C, Berg KL;

XX MPI; 2001-041427/05.

PT Genomic library for identifying diagnostic and therapeutic  
 compositions, and for identifying virulence factors, regulatory

PT elements and drug targets, comprises Moraxella catarrhalis nucleic  
 acids -  
 XX  
 PS Claim 1; Page 293-308; 545pp; English.

CC The present invention relates to a Moraxella catarrhalis genomic library  
 CC comprising of a combination of 41 nucleic acid molecules (see  
 CC AAF28514-AAF28544). The library has a number of uses described in the  
 CC specification e.g. is useful for identifying diagnostic and therapeutic  
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large  
 CC aerobic, gram-negative diplococcus, normally found among the bacterial  
 CC flora of human upper airways. M. catarrhalis is known to cause acute,  
 CC localised infections such as otitis media, sinusitis and bronchopulmonary  
 CC infection and life-threatening, systemic diseases including endocarditis  
 CC and meningitis.

SO Sequence 62909 BP; 18449 A; 13798 C; 12744 G; 17918 T; 0 other;

Query Match 7.5%; Score 45; DB 22; Length 62909;  
 Best Local Similarity 47.7%; Pred. No. 0.011;  
 Matches 132; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

OY 202 aacaagctgactcgttctgaagatcgcccaagaagctcaacctgaccagagctcagcc 261  
 |||||  
 DB 53001 AATTAAGCTCTAATTAAGCTCTAATTAAGCTCTAATTAAGCTCTAATTAAGCTCT 52942  
 |||||  
 OY 262 gagcgtcgttgaagcgttcacagatggtgtcgtcgaagctagaagctcgcggaagc 321  
 |||||  
 DB 52941 GCCAAGCGGTTAATGATTTACTGAAGCTGTCAAGCTGCTTACAGCTGCGGATGAT 52882  
 |||||  
 OY 322 ctgaagctcagcgcgtctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 381  
 |||||  
 DB 52881 GTCGTAATGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 52822  
 |||||  
 OY 382 aaccgcgcacgtcgc 441  
 |||||  
 DB 52821 AATTCAAAAGCTGCGCAAGCAATTCAGCAAGCTGCGCAAGCTGCGCAAGCTGCG 52762  
 |||||  
 OY 442 ggcctccgtcgtcgaagaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 478  
 |||||  
 DB 52761 GGTAAAGTCTTAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 52725  
 |||||

RESULT 6

AAV99811  
 ID AAV99811 standard; DNA; 16075 BP.

AC AAV99811;

DT 14-JUN-1999 (first entry)

DE Gum gene cluster.

XX Xanthan gum; gum; exopolysaccharide; gum gene cluster;

KW recombinant; Xanthomonas; Sphingomonas; sphingans 588; sugar;

OS Xanthomonas campestris.

PN WO9856942-A1.

PF 17-DEC-1998.

XX 12-JUN-1998; 98MO-US12322.

XX 12-JUN-1998; 98US-0096942.

PR 12-JUN-1997; 97US-0049428.

PR 11-JUN-1998; 98US-0096867.

PA (SHIN-) SHINETSU BIO INC.  
 XX (SHIE ) SHINETSU CHEM CO LTD.











PS Disclosure: Page 100-102; 187pp; Japanese.

CC The present invention describes cells originating in bone marrow or  
 CC umbilical blood cells which are capable of differentiating into  
 CC cardiomyocytes. Also described are: (1) cardiomyocytes produced by the  
 CC differentiation of the cells; (2) a method for carrying out the  
 CC differentiation into cardiomyocytes, regulated by a promotional and/or  
 CC inhibitory factor; (3) a method for the differentiation of the cells  
 CC into cell types other than cardiomyocytes; (4) drug compositions  
 CC promoting the formation of heart muscle and regeneration of heart tissue  
 CC which contain the cells; (5) a method for the production of antibodies  
 CC which recognise the cells; especially antibodies which recognise a  
 CC surface antigen on the cells; (6) a method for screening factors which  
 CC promote the proliferation of the cells; (7) a method for immortalising  
 CC the cells by expressing telomerase in them; (8) drug compositions for  
 CC the treatment of heart disease which contain the immortalised cells; and  
 CC (9) cell-free supernatant from the culture of the cells and its use in  
 CC promoting their differentiation into cardiomyocytes. The cells are used  
 CC in the treatment of diseases involving heart muscle degeneration, such  
 CC as myocardial infarction and in the study of cardiomyocyte  
 CC differentiation. AAH44351 to AAH44409 and AAB99935 represent  
 CC sequences used in the exemplification of the present invention.

XX Sequence 1326 BP; 215 A; 499 C; 414 G; 198 T; 0 other;

Query Match 6.5%; Score 39; DB 22; Length 1326;

Best Local Similarity 56.4%; Pred. No. 0.12;

Matches 92; Conservative 0; Mismatches 70; Indels 1; Gaps 1;

QY 235 aagtcacactgacccaaggtcagcgagcgtctgttaacgcttcagagatgtgttc 294

Db 566 AGGACGGGGCTGTCTGAAAGGGGCGCGGAGGCGGCGCTGCGGCGCAGAGCGCGCC 507

QY 295 gtcgaggtatgaagtcgagcgagcgtgaagctcacgcgctgttcgcgctgagcgc 354

Db 506 ACGTCGGCCATGTAAAGCCGGGTAGGGGCTGGAG-TAGGAGCCCGCAGACCGCGCGCC 448

QY 355 gtcaagcgccggtcgcagcagcgccgaacccgcgcgactgagcg 397

Db 447 GTACTGCTCGCGCGCCGCCAGCGCCGACCCCGCTCGCGCG 405

RESULT 15

AAH48225/c

ID AAH48225 standard; DNA; 1326 BP.

AC AAH48225;

DT 21-SEP-2001 (first entry)

DE Heart muscle cell differentiation related DNA SEQ ID NO: 12.

KW Heart muscle cell; human; cell differentiation; heart disease; ds.

OS Homo sapiens.

PN WO200148151-A1.

PD 05-JUL-2001.

PF 27-DEC-2000; 2000WO-JP09323.

PR 28-DEC-1999; 99JP-0372826.

PR 28-FEB-2000; 2000WO-JP01148.

PR 02-NOV-2000; 2000WO-JP07741.

PA (KYO) KYOWA HAKKO KOGYO KK.

PI Umezawa A, Hata J, Fukuda K, Ogawa S, Sakurada K, Gojo S;

PI Yamada Y;

XX WPI; 2001-425656/45.

DR P-PSDB; AAG64849.

XX Cells capable of differentiating into cardiomyocytes and originating in  
 PT bone marrow or umbilical blood cells for study of cardiomyocyte  
 PT differentiation and treatment of heart disease

PS Disclosure: Page 106-108; 183pp; Japanese.

CC The present invention provides cells originating in the human bone marrow  
 CC or umbilical blood cells which are capable of differentiating into  
 CC cardiomyocytes. These cells are useful in the treatment of diseases  
 CC involving heart muscle degeneration, such as myocardial infarction, and  
 CC the study of cardiomyocyte differentiation. The present sequence is  
 CC an oligonucleotide described in the exemplification of the invention.

XX Sequence 1326 BP; 215 A; 499 C; 414 G; 198 T; 0 other;

Query Match 6.5%; Score 39; DB 22; Length 1326;

Best Local Similarity 56.4%; Pred. No. 0.12;

Matches 92; Conservative 0; Mismatches 70; Indels 1; Gaps 1;

QY 235 aagtcacactgacccaaggtcagcgagcgtctgttaacgcttcagagatgtgttc 294

Db 566 AGGACGGGGCTGTCTGAAAGGGGCGCGGAGGCGGCGCTGCGGCGCAGAGCGCGCC 507

QY 295 gtcgaggtatgaagtcgagcgagcgtgaagctcacgcgctgttcgcgctgagcgc 354

Db 506 ACGTCGGCCATGTAAAGCCGGGTAGGGGCTGGAG-TAGGAGCCCGCAGACCGCGCGCC 448

QY 355 gtcaagcgccggtcgcagcagcgccgaacccgcgcgactgagcg 397

Db 447 GTACTGCTCGCGCGCCGCCAGCGCCGACCCCGCTCGCGCG 405

Search completed: December 20, 2001, 20:02:58  
 Job time: 4094 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2001, 18:51:34 ; Search time 88.01 Seconds  
(without alignments)  
1543.993 Million cell updates/sec

Title: US-09-816-391A-1

Perfect score: 1 gctggcgcgcgcgcatga.....tattcgactagtcagcg 600

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents.NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCtUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43.6	7.3	16075	3	US-09-096-942-1
2	43.6	7.3	16075	3	US-09-096-867-1
3	37.6	6.3	858	4	US-09-095-855-200
4	37	6.2	2943	1	US-08-042-747A-7
5	35	5.8	1950	3	US-08-911-853-14
6	35	5.8	1950	4	US-09-479-409-14
7	35	5.8	17612	3	US-08-911-853-29
8	35	5.8	17612	4	US-09-479-409-29
9	35	5.8	4403765	4	US-09-103-840A-2
10	35	5.8	4411529	4	US-09-103-840A-1
11	34.8	5.8	2721	6	5215881-2
12	34.8	5.8	8438	1	US-07-945-283-1
13	34.6	5.7	1524	2	US-08-426-359B-3
14	34.4	5.7	1524	6	5512669-1
15	34.4	5.7	6854	4	US-09-194-905-7
16	34	5.7	8323	1	US-08-110-300A-8
17	34	5.7	8323	2	US-08-886-642-8
18	34	5.7	8323	5	PCT-US93-08041-8
19	34	5.7	10367	2	US-08-110-300A-9
20	34	5.7	10367	5	PCT-US93-08041-9
21	34	5.6	1710	2	US-08-912-794-1
22	33.8	5.6	6379	1	US-08-499-215-1
23	33.8	5.6	6379	2	US-08-997-080-191
24	33.2	5.5	337	2	US-08-997-362-191
25	33.2	5.5	337	2	US-08-997-362-191
26	33.2	5.5	337	2	US-08-997-362-191
27	33.2	5.5	337	4	US-09-095-855-191

C	28	33	5.5	80161	3	US-09-036-987A-1	Sequence 1, Appli
C	29	33	5.5	891	4	US-09-370-700-1	Sequence 1, Appli
C	30	32.8	5.5	12588	2	US-09-049-672A-23	Sequence 23, Appli
C	31	32.8	5.5	1521	1	US-08-387-942C-1	Sequence 1, Appli
C	32	32.4	5.4	1521	1	US-08-083-948-14	Sequence 14, Appli
C	33	32.4	5.4	1521	1	US-08-393-785-14	Sequence 14, Appli
C	34	32.4	5.4	1521	1	US-08-475-694-14	Sequence 14, Appli
C	35	32.4	5.4	1521	1	US-08-712-057-14	Sequence 14, Appli
C	36	32.4	5.4	1647	1	US-08-083-948-13	Sequence 13, Appli
C	37	32.4	5.4	1647	1	US-08-393-785-13	Sequence 13, Appli
C	38	32.4	5.4	1647	1	US-08-475-694-13	Sequence 13, Appli
C	39	32.4	5.4	1647	1	US-08-712-057-13	Sequence 13, Appli
C	40	32.4	5.4	1865	1	US-08-083-948-7	Sequence 7, Appli
C	41	32.4	5.4	1865	1	US-08-393-785-7	Sequence 7, Appli
C	42	32.4	5.4	1865	1	US-08-475-694-7	Sequence 7, Appli
C	43	32.4	5.4	1865	1	US-08-712-057-7	Sequence 7, Appli
C	44	32.2	5.4	15872	4	US-09-105-537-1	Sequence 1, Appli
C	45	32.2	5.4	36741	4	US-09-301-665-3	Sequence 3, Appli

#### ALIGNMENTS

```

RESULT 1
US-09-096-942-1
: Sequence 1, Application US/09096942
: Patent No. 6027925
: GENERAL INFORMATION:
: APPLICANT: Pollock, Thomas J
: APPLICANT: Mikolajczak, Marcia
: APPLICANT: Yamazaki, Motohide
: APPLICANT: Thorne, Linda
: APPLICANT: Armentrout, Richard W
: TITLE OF INVENTION: Production of Xanthan Gum by Sphingomonas Bacteria
: TITLE OF INVENTION: Carrying Genes from Xanthomonas Campestris
: FILE REFERENCE: seq list for appli filed from pro. appli
: CURRENT APPLICATION NUMBER: US/09/096,942
: EARLIER FILING DATE: 1998-06-12
: EARLIER APPLICATION NUMBER: 60/049,428
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 16075
: TYPE: DNA
: ORGANISM: Xanthomonas campestris
: PUBLICATION INFORMATION:
: US-09-096-942-1

```

Query Match 7.3%, Score 43.6; DB 3; Length 16075;  
Best Local Similarity 47.7%; Pred. No. 0.0017;  
Matches 127; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

OY	123	atggatacaaacagctcgtactcgttcgaaagatccgcgaagatccaactgcgaag	252
DB	33	atggatctgaagaagcggagatggtccgagcgctcgttcgaaagatcgtcgaagcag	92
OY	253	gtccagcgcgaagctcgttcaacgccttccaagatggttcgttcgagcgatgaagtc	312
DB	93	ctgtgagcgaaggaattcgttcgacgcgttttcgtatgctcgtcgcgagatcgaagcag	152
OY	313	ggcgaagcctgaagctacccgcctgttcctcgttcgctgagcgcgtlcaagcgcgcgtcgc	372
DB	153	ggccctcaggtgaagttgtcgtcgttcgcgcaattcgtatcgtcgcgcgaagaaccaagc	212
OY	373	acggcgcgaacccgcgcgacgtcgcgcgagatgttgcattcgcgttccttaagcgcttct	432
DB	213	ccggctcgaatcccaagacccggtgaagaaattccgattcgtcgcgaagcagcgttgtgacc	272
OY	433	atctcgcgtcctcctcgtcgaagaa	458
DB	273	ttccgcgcgcgcgcgaactcaagga	298

RESULT 2  
US-09-096-867-1  
Sequence 1, Application US/09096867  
Patent No. 6030817  
GENERAL INFORMATION:  
APPLICANT: Pollock, Thomas J  
APPLICANT: Mikolajczak, Marcia  
APPLICANT: Yamazaki, Motohide  
APPLICANT: Thorne, Linda  
APPLICANT: Armentrout, Richard W  
TITLE OF INVENTION: Production of Xanthan Gum by *Sphingomonas Bacteria*  
TITLE OF INVENTION: Carrying Genes from *Xanthomonas Campestris*  
FILE REFERENCE: seq list for appl filed from pro. appl  
CURRENT APPLICATION NUMBER: US/09/096,867  
CURRENT FILING DATE: 1998-06-11  
EARLIER APPLICATION NUMBER: 60/049,428  
EARLIER FILING DATE: 1997-06-12  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 1  
LENGTH: 16075  
TYPE: DNA  
ORGANISM: *Xanthomonas campestris*  
PUBLICATION INFORMATION:  
US-09-096-867-1

Query Match 7.3%, Score 43.6; DB 3; Length 16075;  
Best Local Similarity 47.7%; Pred. No. 0.0017;  
Matches 127; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

```
OY 193 atggatatacaacagctcgttccgaagatcgcccaagaatcccaactgacaaag 252
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 33 atgcatgaagaaagcggaagatgcccgaagcgtctgttcgaagaagtcggctgaacaa 92
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 253 gctcagcgcgaagctgctgttaacgcctccagaatgtgttcgtagagctgaagtc 312
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 93 cgtgagcggaaggaattcgtcgaacggttttcgattgtctgcgactgacgtgagcag 152
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 313 ggcgaagcgaagcctacccagcgtgttctcgcgtgagcggtcaagcggccgctgc 372
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 153 ggcgcgcagtgagtgctgtcggcttcgcaacttcgactcgcgcgcaagaacacag 212
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 373 accgcgcacacccgcgcactgtagcagacattccgacttcctacgcgttcgt 432
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 213 ccggtcgaatcccaagacgcggtgagaaattccgactcgcgcagagacggtgtgacc 272
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 433 atctcgcgtgctccctcgtgaaga 458
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 273 ttccgccccgcgcagaactcaaga 298
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 3  
US-09-095-855-200  
Sequence 200, Application US/09095855  
Patent No. 6160093  
GENERAL INFORMATION:  
APPLICANT: Tan, Paul  
APPLICANT: Visser, Elizabeth  
APPLICANT: Skinner, Margot  
APPLICANT: Prestidge, Ross  
TITLE OF INVENTION: Compounds and Methods for *Mycobacterial Infections*  
TITLE OF INVENTION: Treatment and Diagnosis of *Mycobacterial Infections*  
NUMBER OF SEQUENCES: 208  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: law offices of Ann W. Spectman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/095,855  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/705,347  
FILING DATE: 29-AUG-1996  
APPLICATION NUMBER: 08/873,970  
FILING DATE: 12-JUN-1997  
APPLICATION NUMBER: 08/997,362  
FILING DATE: 23-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Sleath, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000,1002c3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563  
TELEX:  
INFORMATION FOR SEQ ID NO: 200:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 858 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-09-095-855-200

Query Match 6.3%, Score 37.6; DB 4; Length 858;  
Best Local Similarity 46.3%; Pred. No. 0.027;  
Matches 124; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

```
OY 202 aacaaagtgcctcgtttcgaagatcgcccaagaatcccaactgacaaagctcgaagc 261
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 187 AACAAAGCAGAGCTCATGACGTACTGAGAGCTGGCTCGATCGATCGGCAAGCG 246
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 262 gagctgtgttaagccttcacgattgttctcgcgtgagcggtatagaatccggcgaaagc 321
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 247 ACTGGGCGGTGGAAGAGTGTGCAACCAATCGTGCCTGCCGCTGCACAAAGGTGAGAGC 306
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 322 ctgaagctcacccgctgttctcgcgtgagcggtcaagcggccgctcgaacggccgc 381
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 307 GTCACATCAGCGGCTTGATGTTTTCAGAGCAGCGTCTCGCGCAGCAGCGGTGCACAGC 366
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 382 aaccgcgaactggcgagatgacattccgacttcctacagcggttcgtatctcgcgt 441
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 367 AATCCGCGCAGCGGAGAACCGTGAAGGTCAAGCCACCTCAGTCCGCGATTCGCTCC 426
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 442 ggtcctcgtcgaagaagcgctcacgc 469
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 427 GCGCTCAGTTCAAGGCTGTGTCTG 454
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 4  
US-08-042-747A-7  
Sequence 7, Application US/08042747A  
Patent No. 5487969  
GENERAL INFORMATION:  
APPLICANT: Eberle, Richard  
APPLICANT: Black, Darla  
APPLICANT: Scintcarillo, Franco  
APPLICANT: Hilliard, Julia K.  
TITLE OF INVENTION: Cloning and Amplification of Monkey B  
TITLE OF INVENTION: Virus Genes  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cox & Smith Incorporated

STREET: 112 East Pecan Street, Suite 2000  
CITY: San Antonio  
STATE: Texas  
COUNTRY: USA  
ZIP: 78205  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/042,747A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Haymond, W. Bradley  
REGISTRATION NUMBER: 35186  
REFERENCE/DOCKET NUMBER: S-0072.179  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 210-554-5500  
TELEFAX: 210-226-8395  
TELEX: 767609  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2943 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 87..2744  
US-08-042-747A-7

Query Match 6.2%; Score 37; DB 1; Length 2943;  
Best Local Similarity 47.6%; Pred. No. 0.081;  
Matches 109; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 200 acaaaagctgctgacctgcttcgaagatcgcccaagagtcacacccgacgaagctcagc 259  
DB 2452 ACCCATGAGAGCCCTGTCACCGCTACACCAAGAGAGCTCAAGAGAGCGAGCGCCCGC 2511  
QY 260 ccgaagctgctgttaagcgttcacagatgltgttcgtcgagctatgaagctccgcgcaag 319  
DB 2512 TGGCGGGCGCGGCGAGGAGCGCGCGAGACTTTTGACGAGCGCAAGCGGCGCAGCGCG 2571  
QY 320 gctgaagctacccgctgctgtctccgctgagcgcgctcaagcgcccgctcgacccgac 379  
DB 2572 GGGAGATGATCGCTACATGAGCGCTGCTCGGCCATGAGAGCGCACCGGACCAAGGCC 2631  
QY 380 gcaaccgcgagcactgagcgagcagattgacattccgagcttcctcctgaagcgt 428  
DB 2632 GCAAGAAGGCGACGAGCGCCCTGCTGAGCGCGAAGGTTACCGAGCGCGGT 2680

RESULT 5  
US-08-911-853-14  
Sequence 14, Application US/08911853  
Patent No. 6048710  
GENERAL INFORMATION:  
APPLICANT: Geritise, Gijshert  
APPLICANT: Quax, Wilhelmus J.  
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESS: Genencor International  
STREET: 925 Page Mill Road  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1013

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/911,853  
FILING DATE:  
Prior Application Data:  
APPLICATION NUMBER: 08/699,092  
FILING DATE: 16-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gialster, Debra J  
REGISTRATION NUMBER: 33,888  
REFERENCE/DOCKET NUMBER: GC361-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-846-7620  
TELEFAX: 650-845-6504  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1950 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-911-853-14

Query Match 5.8%; Score 35; DB 3; Length 1950;  
Best Local Similarity 50.3%; Pred. No. 0.28;  
Matches 86; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 209 ctgacctgcttcgaagatcgcccaagagtcacacccgacgaagctcagcagcgagcgtg 268  
DB 572 CCGAGCTCCATGCCAGAGCGCGGCGGAGCTACACCTGATCACTGCGACATGCTGGG 631  
QY 269 ctgttaagccttcacagatgltgttcgtcagagctatgaagctcgagcgagcctgaagc 328  
DB 632 TACTGAGCGCCCGCGAGCGAGCTGAACACGCGGTGATGCGCAAGGAAAAACAGCGCG 691  
QY 329 taaccgagctgttcctccgtgagcggttaagcgcccgctcgacccgac 379  
DB 692 GCACCGGGGTATGCGCGAGCGCCGCGACCAACCGCGCTGATCCTCGCGCC 742

RESULT 6  
US-09-479-409-14  
Sequence 14, Application US/09479409  
Patent No. 6225106  
GENERAL INFORMATION:  
APPLICANT: Geritise, Gijshert  
APPLICANT: Quax, Wilhelmus J.  
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESS: Genencor International  
STREET: 925 Page Mill Road  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1013  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/479,409  
FILING DATE:  
Prior Application Data:  
APPLICATION NUMBER: 08/911,853  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:









```

:
: FILING DATE: 03-NOV-1992
: APPLICATION NUMBER: 566,535
: FILING DATE: 13-AUG-1990
: APPLICATION NUMBER: 67,695
: FILING DATE: 29-JUN-1987
:
: SEQ ID NO.: 1
:
: LENGTH: 1524
5512669-1

```

Query Match	5.7%;	Score 34.4;	DB 6;	Length 1524;
Best Local Similarity	49.48;	Pred. No. 0.37;		
Matches	89;	Conservative	0;	Mismatches 91;
			Indels	0;
			Gaps	0;

QY	201	caacaaagctcaactctgttcgaaagatcgcacccaaagtcacaactcgacaaagctcaagc	260
Db	359	caatacgcgcgcgcacataaactcggcgacgcggatatttcggcggttctcgaagcgcgcggg	418
QY	261	cgaagctctgtttaaagcccttcacagagatgtgtctcgtcgaagctcagatgcgcgcgaag	320
Db	419	cgttcgcgcgcgcgcgaagtgaacgcgcgcgcgtatctctcgcgcacagctgcctgcgcgcgcgaag	478
QY	321	cctgaagctcaacgcgcgcgtgttctcctcgcgtcgaagcgcgcgcgcgcctcgaacgcgcgc	380
Db	479	ccagaaaccgcgcgcgcgcgaagcgcgcgcacagaaagccgcgcgtcgcgcgaagcgaagcgaacgcctg	538

```

RESULT 15
5229279-1
:
: Patent No. 5229279
:
: APPLICANT: PEOPLES, OLIVER P.; SINSKEY, ANTHONY J.
:
: TITLE OF INVENTION: METHOD FOR PRODUCING NOVEL POLYESTER
: BIOPOLYMERS
:
: NUMBER OF SEQUENCES: 8
:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER: US/07/5556,535
:
: FILING DATE: 13-AUG-1990
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: 67,695
:
: FILING DATE: 29-AUG-1987
:
: SEQ ID NO:1:
:
: LENGTH: 1525
:
: 5229279-1

```

Query Match	5.7%	Score 34.4;	DB 6;	Length 1525;
Best Local Similarity	49.4%;	Pred. No. 0.37;		
Matches	89;	Conservative	0;	Mismatches 91;
			Indels	0;
			Gaps	0;

QY	Db	QY	Db
201	caacaaatctbactcgtttcgaagatcgcgccaaagatccaaactgcacgaagctcaagc	261	cgagagcgcgtcttaacgccttcacagatgctgtctcgaagcatatgaagcgcgcgaagc
202		262	
203		263	
204		264	
205		265	
206	caacacgcccgcgccatgaactcggcgagccggtattctcgcggtatctcgaagcgcgcgagc	266	cgatgagcgcgtcttcaacgccttcacagatgctgtctcgaagcatatgaagcgcgcgaagc
207		267	
208		268	
209		269	
210		270	
211		271	
212		272	
213		273	
214		274	
215		275	
216		276	
217		277	
218		278	
219		279	
220		280	
221		281	
222		282	
223		283	
224		284	
225		285	
226		286	
227		287	
228		288	
229		289	
230		290	
231		291	
232		292	
233		293	
234		294	
235		295	
236		296	
237		297	
238		298	
239		299	
240		300	
241		301	
242		302	
243		303	
244		304	
245		305	
246		306	
247		307	
248		308	
249		309	
250		310	
251		311	
252		312	
253		313	
254		314	
255		315	
256		316	
257		317	
258		318	
259		319	
260		320	
261		321	
262		322	
263		323	
264		324	
265		325	
266		326	
267		327	
268		328	
269		329	
270		330	
271		331	
272		332	
273		333	
274		334	
275		335	
276		336	
277		337	
278		338	
279		339	
280		340	
281		341	
282		342	
283		343	
284		344	
285		345	
286		346	
287		347	
288		348	
289		349	
290		350	
291		351	
292		352	
293		353	
294		354	
295		355	
296		356</	

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2001, 17:33:44 ; Search time 1563.06 seconds  
(without alignments)  
4124.901 Million cell updates/sec

Title: US-09-816-391A-1

Perfect score: 600  
Sequence: 1 gctggcgcgcgccatgta.....tattccgactagtcacgcg 600

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_estfun:\*  
3: em\_estfun:\*  
4: em\_estfun:\*  
5: em\_estfun:\*  
6: em\_estfun:\*  
7: em\_estfun:\*  
8: em\_estfun:\*  
9: em\_estfun:\*  
10: em\_estfun:\*  
11: em\_estfun:\*  
12: em\_estfun:\*  
13: em\_estfun:\*  
14: em\_estfun:\*  
15: em\_estfun:\*  
16: em\_estfun:\*  
17: em\_estfun:\*  
18: em\_estfun:\*  
19: em\_estfun:\*  
20: em\_estfun:\*  
21: em\_estfun:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	59.6	9.9	428	10	AV386687	AV386687 AV386687
2	59.6	9.9	447	10	AV393637	AV393637 AV393637
3	59.6	9.9	465	10	AV634084	AV634084 AV634084
4	59.6	9.9	509	10	AV387318	AV387318 AV387318
5	59.6	9.9	625	10	BE337777	BE337777 BE337777
6	59.6	9.9	696	11	BG860650	BG860650 BG860650
7	59.4	9.9	592	13	AO990760	AO990760 AO990760
8	55	9.2	376	10	AV637329	AV637329 AV637329
9	51.6	8.6	627	13	AO989976	AO989976 AO989976
10	49.8	8.3	800	13	AZ934085	AZ934085 AZ934085
11	46.2	7.7	663	11	BF483682	BF483682 BF483682
12	44	7.3	831	11	BG309663	BG309663 BG309663

13	43.6	7.3	1201	13	CNS016BR	AL106545 Drosophila
14	42.6	7.1	493	10	BE356624	BE356624 Drosophila
15	42.6	7.1	578	11	BI099174	BI099174 Drosophila
16	42.6	7.1	651	11	BG355574	BG355574 Drosophila
17	42.4	7.1	761	13	AO991749	AO991749 Drosophila
18	41.8	7.0	458	10	AM679463	AM679463 Drosophila
19	41.6	6.9	925	13	CNS0094P	AL053013 Drosophila
20	41	6.8	509	11	BE485321	BE485321 Drosophila
21	40.6	6.8	569	11	BE194806	BE194806 Drosophila
22	40.4	6.7	785	10	BE356631	BE356631 Drosophila
23	40	6.7	304	10	BE761429	BE761429 Drosophila
24	39.8	6.6	304	13	AZ578311	AZ578311 Drosophila
25	39.8	6.6	550	10	AM678194	AM678194 Drosophila
26	39.6	6.6	529	10	BE485321	BE485321 Drosophila
27	39.6	6.6	1036	13	CNS010BS	AL053013 Drosophila
28	39.4	6.6	435	10	AM502914	AM502914 Drosophila
29	39.4	6.6	574	11	BE597133	BE597133 Drosophila
30	39.2	6.5	594	10	BE361820	BE361820 Drosophila
31	39	6.5	396	10	AM293155	AM293155 Drosophila
32	39	6.5	418	11	BE058235	BE058235 Drosophila
33	39	6.5	434	10	AL187315	AL187315 Drosophila
34	39	6.5	440	10	AL195171	AL195171 Drosophila
35	39	6.5	495	10	AL1218830	AL1218830 Drosophila
36	39	6.5	495	11	BF058124	BF058124 Drosophila
37	39	6.5	500	11	BE514120	BE514120 Drosophila
38	39	6.5	517	10	AL1218820	AL1218820 Drosophila
39	39	6.5	519	10	AL052253	AL052253 Drosophila
40	39	6.5	535	11	BE513628	BE513628 Drosophila
41	39	6.5	557	10	AM027166	AM027166 Drosophila
42	39	6.5	603	11	BE513289	BE513289 Drosophila
43	39	6.5	606	11	BF448729	BF448729 Drosophila
44	38.8	6.5	617	13	AZ934220	AZ934220 Drosophila
45	38.4	6.4	423	10	AO083552	AO083552 Drosophila

#### ALIGNMENTS

RESULT 1  
LOCUS AV386687 428 bp mRNA  
DEFINITION AV386687 Chlamydomonas reinhardtii c9 Chlamydomonas reinhardtii  
CDNA clone CM004f12\_r, mRNA sequence.  
ACCESSION AV386687  
VERSION AV386687.1 GI:6540903  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii.  
ORGANISM Chlamydomonas reinhardtii.  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadaceae; Chlamydomonas.  
1 (bases 1 to 428)  
Assamizu,E., Nakamura,Y., Sato,S., Fukuzawa,H. and Tabata,S.  
A large scale structural analysis of cDNAs in a unicellular green  
alga, Chlamydomonas reinhardtii. I. Generation of 3433  
non-redundant expressed sequence tags  
JOURNAL DNA Res. 6 (6), 369-373 (1999)  
MEDLINE 20152986  
COMMENT Contact: Yasukazu Nakamura  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: ynakamu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/  
Location/Qualifiers  
1. 428  
/organism="Chlamydomonas reinhardtii"  
/strain="C9"  
/db\_xref="taxon:3055"  
/clone="CM004f12\_r"  
/clone\_id="Chlamydomonas reinhardtii c9"  
/dev\_stage="photoautotrophic growth"  
/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
XhoI"  
BASE COUNT 73 a 146 c 140 g 69 t



OY 384 cccgcgcactgagcgagcaatgacattccggtctctacgagcggttcgtatctccgctg 443  
 DB 133 CCCGACGACGGCGCGCTCCAGATCCCGCCAGCAGGCCCCGTTCAAGGCTTC 242  
 OY 444 ctccctgtcgaagaagcgccgtcacg 469  
 DB 243 AGTTGGCCTGAAGATGCCGTGAACG 268

RESULT 4  
 AV387318 509 bp mRNA EST 29-SEP-2000  
 LOCUS AV387318 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii

DEFINITION CDNA clone CM019f02\_r, mRNA sequence.  
 AV387318  
 AV387318.1 GI:6541534

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

REFERENCE  
 AUTHORS  
 TITLE

A large scale structural analysis of cDNAs in a unicellular green non-redundant expressed sequence tags

JOURNAL  
 MEDLINE  
 COMMENT

Contact: Yasukazu Nakamura  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: ynakamura@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
 source

1. 509  
 /organism="Chlamydomonas reinhardtii"  
 /strain="C9"  
 /db\_xref="taxon:3055"  
 /clone\_lbp="CM019f02\_r"  
 /clone\_lbp="Chlamydomonas reinhardtii C9"  
 /dev\_stage="photoautotrophic growth"  
 /note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2: XhoI"  
 BASE COUNT 82 a 182 c 168 g 77 t  
 ORIGIN

Query Match 9.9%; Score 59.6; DB 10; Length 509;  
 Best Local Similarity 51.5%; Pred. No. 1.2e-06;  
 Matches 137; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

OY 204 caagctgacctgcttgcgaagatcgccagaagtcacactgacgaagtcagcgca 263  
 DB 125 CAAGGCTAAACTTGTGAGAGCGATCGCTACTGAGCTGGGACTGACCAAGAGCGTGGC 184  
 OY 264 gctgtctgttaacgcttccagagatggttcgttcgaagtcatagaagtcagcgca 323  
 DB 185 GAAGGCTTTGACTCCCTGATGTTGGCGGCAATTGAGAGCGCCCTCATCAACGGTGAACCCGCT 244  
 OY 324 gaagctacacgagctgttctccgctgagcgctcaagcgccgagtcgacacgagcgca 383  
 DB 245 GACCATCTGTGGGCTTGGGCAAGTTGAGTGCAGCGGCGCTGCGGCAAGGCGCGCA 304  
 OY 384 ccgcgcactgagcgagcaagatgacattcggttctcctaagggttcgtatctccgctg 443  
 DB 305 CCCGACGACGGCGCGCTCTCCAGATCCCGCCAGCAGGCCCCGTTCAAGGCTTC 364  
 OY 444 ctccctgtcgaagaagcgccgtcacg 469  
 DB 365 AGTTGGCCTGAAGATGCCGTGAACG 390

RESULT 5

BE337777 625 bp mRNA EST 14-JUL-2000  
 LOCUS BE337777 Chlamydomonas reinhardtii CC-1690, normalized, lambda Zap II  
 DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

REFERENCE  
 AUTHORS  
 TITLE

1 (bases 1 to 625)  
 Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P.,  
 McDermott, J. P., Silflow, C., Stern, D. and Surzycki, R.  
 Analyses of the Chlamydomonas reinhardtii Genome: A Model,  
 Unicellular System for Analyzing Gene Function and Regulation in  
 Vascular Plants; project phase 2

JOURNAL  
 MEDLINE  
 COMMENT

Contact: Elizabeth H. Harris  
 Duke University  
 Durham, NC 27708-1000, USA  
 Tel.: 919 613 8164  
 Fax: 919 613 8177  
 Email: chlamy@duke.edu.

FEATURES  
 source

1. 625  
 /organism="Chlamydomonas reinhardtii"  
 /strain="CC-1690 wild type mt+ 21gr"  
 /db\_xref="taxon:3055"  
 /clone\_lbp="C. reinhardtii CC-1690, normalized, Lambda Zap II"  
 /note="Vector: pBluescript II SK-; Site\_1: EcoRI; Site\_2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with Exsist (Stratagene) phage. The library was normalized using method 4 described in Bonaldi et al (1996) Genome Research 6: 791-806."

BASE COUNT 106 a 216 c 198 g 104 t  
 ORIGIN

Query Match 9.9%; Score 59.6; DB 10; Length 625;  
 Best Local Similarity 51.5%; Pred. No. 1.2e-06;  
 Matches 137; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

OY 204 caagctgacctgcttgcgaagatcgccagaagtcacactgacgaagtcagcgca 263  
 DB 154 CAAGGCTAAACTTGTGAGAGCGATCGCTACTGAGCTGGGACTGACCAAGAGCGTGGC 213  
 OY 264 gctgtctgttaacgcttccagagatggttcgttcgaagtcatagaagtcagcgca 323  
 DB 214 GAAGGCTTTGACTCCCTGATGTTGGCGGCAATTGAGAGCGCCCTCATCAACGGTGAACCCGCT 273  
 OY 324 gaagctacacgagctgttctccgctgagcgctcaagcgccgagtcgacacgagcgca 383  
 DB 274 GACCATCTGTGGGCTTGGGCAAGTTGAGTGCAGCGGCGCTGCGGCAAGGCGCGCA 333  
 OY 384 ccgcgcactgagcgagcaagatgacattcggttctcctaagggttcgtatctccgctg 443  
 DB 334 CCCGACGACGGCGCGCTCTCCAGATCCCGCCAGCAGGCCCCGTTCAAGGCTTC 393  
 OY 444 ctccctgtcgaagaagcgccgtcacg 469  
 DB 394 AGTTGGCCTGAAGATGCCGTGAACG 419

RESULT	6
LOCUS	BS860650
DEFINITION	1024072A08.y1 C. reinhardtii CC-1690, normalized, lambda zap II
ACCESSION	BS860650
VERSION	BS860650.1
KEYWORDS	GI:14241834
SOURCE	EST.
ORGANISM	Chlamydomonas reinhardtii.
REFERENCE	Chlamydomonas reinhardtii. Eukaryotic Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.
ATTN	1 (bases 1 to 696)
ATTN	Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P., McDermott, J.P., Silflow, C., Stern, D. and Strzyski, R.
ATTN	Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants; project phase 2
JOURNAL	Unpublished (2000)
COMMENT	Contact: Charles Hauser DCMB Box 91000 Duke University Durham, NC 27708-1000 Tel: 919 613 8159 Fax: 919 613 8177 Email: chauser@duke.edu.
FEATURES	Location/Qualifiers
SOURCE	1..696
	/organism="Chlamydomonas reinhardtii"
	/strain="CC-1690 wild type mt+ 21gr"
	/db_xref="taxon:3055"
	/clone_lib="C. reinhardtii CC-1690, normalized, lambda zap II"
	/note="Vector: pluescript II SK-; Site_1: EcoRI; Site_2: XhoI; This library, constructed by John Davies and Jeffrey Mcdermott, combines cDNAs from CC-1690 cells grown to mid-log phase in YAP (acetate-containing) medium in the light, YAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with Exassist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT	120 a 238 c 222 g 116 t
ORIGIN	
Query Match	9.9% Score 59.6; DB 11; Length 696;
Best Local Similarity	51.5% Pred. NO. 1.3e-06;
Matches 137; Conservative	0; Mismatches 129; Indels 0; Gaps 0;
204	caagctgacgtcgtttcgaagatgcgccagaagccaacctgaccgaagctcagccga 263
137	CAAGGCTTAACTGTGGAGGCGCATGCTGCTACGTAGGTGGAGTGCACCAAGAGTGGCTGC 196
264	ggtctgtcgttaacgctccacagatgctgttcgtcgaagctatgaaagctcgcgcgaagcct 323
197	GAAGGCTTTGATCTCCGATGATGGCGGCGCATGGAGGACCCCTCATCAACAGGAGACCGCGT 256
324	gaagtaacccggccgtcttcgcgtcgaagcgcgtcaagggccggctgcacccggcgcaaa 383
257	GACCATCTGTGGGCTTCGGACAGCTTTGAGGTGCGCGAGCGCGCTGCGCGCAAGGCGGCA 316
384	ccgcgcacatcgcgcgaagatgaacattccgcgtcttcacagcgcgttcgatctccgcgtg 443
317	CCCCAGCAGACGGCGGGGCTCTCCAAATGCCCCCAGCAAGGCCCCCGTTCAAAGGCTC 376
444	ctccctgtcgaagaagccgtcaccg 469

Db	377	AGTTGGCCTGAAGATGCCGTGAACG	402
RESULT	7		
LOCUS	AQ990760/c	592 bp	DNA
DEFINITION	RC01581 Photorhabdus luminescens strain W14 M13 library		GSS 14-AUG-2000
ACCESSION	Photorhabdus luminescens genomic clone PLG01581, DNA sequence.		
VERSION	AQ990760		
KEYWORDS	AQ990760.1 GI:9649354		
SOURCE	GSS.		
ORGANISM	Photorhabdus luminescens.		
	Photorhabdus luminescens		
	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Photorhabdus.		
REFERENCE	1 (bases 1 to 592)		
AUTHORS	french-Constant,R.H., Waterfield,N., Burland,V., Pena,N.T., Daborn,P.J., Bowen,D. and Blattner,F.R.		
TITLE	A genomic sample sequence of the entomopathogenic bacterium Photorhabdus luminescens W14: potential implications for virulence		
JOURNAL	Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)		
MEDLINE	20378633		
COMMENT	Contact: french-Constant RH Department of Biology and Biochemistry University of Bath South Building, Bath BA2 7AY, UK Tel: (44) 1225 826621 Fax: (44) 1225 826779 Email: bsr@bath.ac.uk This is one of 2,122 random reads from the M13 library. For annotation of identified clones (BLASTX, BLASTN and mapping to E. coli K12 genome) please see french-Constant et al. 2000, Nucleic Acids Res. Seq primer: M13 Forward Class: Shotgun. Location/Qualifiers 1..592 /organism="Photorhabdus luminescens" /strain="W14" /db_xref="taxon:29488" /clone="PLG01581" /clone.lib="Photorhabdus luminescens strain W14 M13 library" /dev-stage="primary phase variant" /note="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janus." BASE COUNT 146 a 140 c 112 g 191 t 3 others ORIGIN		
Query Match	9.9%;	Score 59.4;	DB 13; Length 592;
Best Local Similarity	51.3%;	Pred. No. 1.4e-06;	
Matches 138; Conservative	0;	Mismatches 131;	Indels 0; Gaps 0
Db	210	tgacctcgcttcggaatcgcccaagaatccaactgcaagctcagggccgagctgc	269
Db	293	TGAATTAGTATGATCATTCGCAGAAAGCACAGACCTGATTAAGCTCAGCGAAAGCTGC	234
Db	270	tgttaacgcttcacagaatgltctgcgaagctatgaagtcgcgcgaagcgctgaagct	329
Db	233	TCTGGAATCAACTTTGAAATGCATACCCGAAATCCTGGAAGCGGGGATTCAGTCAACT	174
Db	330	caacgcctgtctccgctcgtgagcgctgaagcgccgcgcgcgcgcgcgcgcgcgcgcgcg	389
Db	173	GGTAGGTTTGGGCACTTTCAAAGTTTAACCATTCGTGCAGAACGTAGGGTTCGTAACTCCGA	114
Db	390	cactggagcagacatgacattccggcttcctactacggcgttcgtatctccgctgctccct	449
Db	113	AACCGGTAAGAATATTACATATCGCGGTGGAACAGTACACTGGTTTCTCTGCTGTAAGGC	54
Db	450	gctgaagaaggcgttcaacgagtgacct	478
Db	53	TTTGAAGACGCACTTAAATATTTATTT	25

FEATURES	source
BASE COUNT	64 a 128 c 116 g 68 t
ORIGIN	
Query Match	9.2%; Score 55; DB 10; Length 376;
Best Local Similarity	52.9%; Pred. No. 2,3e-05;
Matches 118; Conservative 0; Mismatches 105; Indels 0; Gaps 0;	
Db	
204	caagctcgacctcgcttcggaagatgccacagaagtcacaactcgaacaaagctcaagccga 263
151	CAAGGCTTAAACTTTGGAGGGGAGATGCTACATGAGGTGGAGATGACCAAGAGAGTGGCTGC 210
Qy	
264	gacgcgtgtaagcgcctccacagatgctgttcgtagagctatgaagctccgcggaagcct 323
Db	
211	GAAAGCCTTAACTCCCGATGATGGCGGCGCATTTGAGACCCCTCATCAACAGGTGACCGCT 270
Qy	
324	gaagctcaacgcgcctgtctctcgcgtgaagcgcgtcaagcgcgcgcgctcgacacgcgcgca 383
Db	
271	GACCATCGTGGGGCTTCCGCGACAGTTTGAAGTGCAGCGAGCGCTGCGCGCCAGGGCGCGAA 330
Qy	
384	ccgcgcgcacgtggcagcagcagatgacattccgcgtctccaaagc 426
Db	
331	CCCCAGCACCGCGGGGTCTCCACAGATGCCGCCAGCAAGAGCC 373
RESULT	
9	
LOCUS	AO989976 627 bp DNA GSS 14-AUG-2000
DEFINITION	Rec00661 Photorhabdus luminescens strain W14 M13 library
ACCESSION	AO989976
VERSION	AO989976.1 GI:9648570
KEYWORDS	GSS.
SOURCE	Photorhabdus luminescens.
ORGANISM	Photorhabdus luminescens.
	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

REFERENCE	
AUTHORS	1 french-Constant,R.H., Waterfield,N., Burland,V., Perna,N.T., Daborn,P.J., Bowen,D. and Blattner,F.R.
TITLE	A genomic sample sequence of the entomopathogenic bacterium Photorhabdus luminescens W14: potential implications for virulence
JOURNAL MEDLINE	Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
COMMENT	Contact: french-Constant RH Department of Biology and Biochemistry University of Bath South Building, Bath BA2 7AY, UK Tel.: (44) 1225 826621 Fax: (44) 1225 826779 Email: bsr@cefbath.ac.uk This is one of 2,122 random reads from the M13 library. For annotation of identified clones (BLASTX, BLASTN and mapping to E.coli K12 genome) please see french-Constant et al. 2000, Nucleic Acids Res. Seq primer: M13 Forward Class: shotgun.
FEATURES	Location/Qualifiers
SOURCE	1..627 /organism="Photorhabdus luminescens" /strain="W14" /db_xref="taxon:29488" /cclone="PIC00661" /clone.lib="Photorhabdus luminescens strain W14 M13 library" /dev_stage="primary phase variant" /note="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janus."
BASE COUNT	189 a     131 c     137 g     166 t     4 others
ORIGIN	
Query Match	8.6%; Score 51.6; DB 13; Length 627;
Best Local Similarity	50.8%; Pred. No. 0.00027;
Matches 123; Conservative	0; Mismatches 119; Indels 0; Gaps 0;
OY	202 aaccagctgcacctggttcgaagtgcgcccaaggatccaactcgacaagctcagcc 261   DB 229 AATAAGACTGATATTGTTCGATGCACATCCGACAAGCAGAACCCTCAAACTTAGCCA 288  OY 262 gagctgttgttaacggccttccagaatgttgctgcgtagcatgaatgcggaagcc 321   DB 289 AAGCGTCTGTGGATCAACTTTGAATGCATCACCGAATCCCTGAAGCGGGTGATTCGA 348  OY 322 ctgaagctcacgcgctgtttcttcgctgtagagcgctcaagcgcccgcgcgcgcgcgcgc 381   DB 349 GTACACACTGGTAAGGTTTTTGGGACTTTTAAAAGTTAACCATCGTGCAGAAACGTCGGTCT 408  OY 382 aaaccgcgcacatggcgagcagattgatcatacttcgcgcttctaagcgcgttcgatatccgcgt 441   DB 409 AACCGCAAACCGGTAAGAATAATACATCCGCCGTGCAAAAGTACCTGTTCTCTCTCT 468  OY 442 gg 443    DB 469 GG 470
RESULT 10	
LOCUS	AZ934085                800 bp        DNA                GSS                24-Apr-2001
DEFINITION	BZ_Be0001D1Br B. japonicum BAC library Bradyrhizobium japonicum
ACCESSION	AZ934085
VERSION	AZ934085.1 GI:13776145
KEYWORDS	GSS.
SOURCE	Bradyrhizobium japonicum. Bradyrhizobium japonicum Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Bradyrhizobium group; Bradyrhizobium.

Query Match	8.3%	Score 49.8	DB 13	Length 800
Best Local Similarity	54.8%	Pred. No. 0.00097		
Matches 121	Conservative	0	Mismatches 97	Indels 3
			Gaps	1
Db 184	ggatgcttattatgcatacaacaagctgacctcgittcogaagatgcc--cagaagtc	240		
Db 388	ggggctcattctctgcattatcattcaatcccaactgtttcagcctatgcgcagacacacccg	447		
Db 241	aactgaccgaagagctcgaagccgctgctgttaacgcttcacgaatgctgtcgaag	300		
Db 448	catctgtracacgacgggaatgtcgagacactttgtgaattgccaatctcgamaagatcgtagcg	507		
Db 301	gcatatgaagtcgagcgaagccttgaagcttcacccgacctgtttccgctgagcgctcaag	360		
Db 508	gctctccgacacccggcgatcgcgtcagctgtggcgccgttcggccttcctcggtcaagcat	567		
Db 361	cgcccgagctcgaccgagccgaccccgagcactgtgcagca	401		
Db 568	cgccctgcacacgcgcggcgctaatccggcgcacccgcgccca	608		
RESULT 11				
LOCUS	BF483682			
DEFINITION	WHE2336.F03.10663 wheat pre-anthesis spike cDNA library Triticum aestivum cDNA clone WHE2336.F03.L06, mRNA sequence.			
ACCESSION	BF483682			
VERSION	BF483682.1			
KEYWORDS	EST.			
SOURCE	Bread wheat.			
ORGANISM	Triticum aestivum			
REFERENCE	Eukaryotes: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae; Pooidaeae; Triticeae; Triticum.			
AUTHORS	1 (bases 1 to 663)			
TITLE	Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Iazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L. and Tong,J.C.			
JOURNAL	The structure and function of the expressed portion of the wheat genomes - Pre-anthesis spike cDNA library			
COMMENT	Unpublished (2000)			
	Contact: Olin Anderson			
	US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center			

```

/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db xref="taxon:4565"
/clone="WHE2336_F03_I06"
/clone_lib="Wheat pre-anthesis spike cDNA library"
/tissue_type="Spike before anthesis"
/dev_stage="Adult plant"
/lab_host="E. coli S0LR"
/note="Vector: Lambda uni-ZAP XR, excised phagemid;
Site.1: EcoRI; Site.2: XhoI; Plants were grown in the
greenhouse. Whole spike with awns trimmed, white, green
and yellow anther were collected and total RNA, and
poly(A) RNA were prepared, a cDNA library was made, and
the cDNA clones were in vivo excised to give phagescript
phagemids in the T7 Close lab (Choi, Close, Penion) at
the University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (all other authors)."

```

RESULT	12
LOCUS	BG309663
DEFINITION	BG309663 831 bp mRNA EST 22-FEB-2001 HVSMEC001M08f Hordeum vulgare seedling shoot EST library HVCNDNA003 (Etioled and unstrressed) Hordeum vulgare cDNA clone HVSMEC001M08f, mRNA sequence.
ACCESSION	BG309663
VERSION	BG309663.1 GI:13110510
KEYWORDS	EST.
SOURCE	barley. Hordeum vulgare
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae Triticeae; Hordeum. 1 (bases 1 to 831)
REFERENCE	Wang, R., Close, T.J., Kleinofcs, A., Wise, R., Begum, D., Frisch, D., Yu, Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo T., Saski, C., Schwartzbeck, J., Simons, J., Choi, D.W., Main, D. and Wood, T.
AUTHORS	Development of a genetically and physically anchored EST resource for barley genomics
TITLE	





Email: mmp@tuega.edu  
Sequences have been trimmed to exclude PolyA, vector and regions below phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: PolyTmX  
High quality sequence stop: 488  
POLYA-No.

FEATURES  
Source Location/Qualifiers

1..493  
/organism="Sorghum bicolor"  
/db\_xref="taxon:4558"  
/clone\_11b="Dark Grown 1 (DG1)"  
/note="Organ: 5-day-old dark-grown seedlings; Vector: lambda zap; Site\_1: XhoI; Site\_2: EcoRI. The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."  
BASE COUNT 78 a 152 c 160 g 103 t  
ORIGIN

Query Match 7.18; Score 42.6; DB 10; Length 493;  
Best Local Similarity 47.9%; Pred. No. 0.1;  
Matches 123; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

194 tggcatacaaaagctcgactcgcttcgaagaatcgcccaagaatccaactgaacaaag 253  
118 TGGGTTCGACCCGAGGCGCGGTGTTCACGCGGAGAGATGACGCGCATGAGACGG 177  
254 ctccgagcgaagctgctgttaacgccttcacagatgtgttcgtagagctatgaatcgcg 313  
178 ACTGGGAAGGTTCAAGCCGCCGCCGCCGAGAGAGAGAGATGAGAGAGAGAGTTCA 237  
314 gcgaaagcctgaagctaacgcgctgttcctccgctgaagcgctgaagcgccgcgcga 373  
238 TGCAGGCTTCACCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 297  
374 ccggccgcaaacccgcgacgagcagagatgagacatcccgcttcctcctgaagcgcttga 433  
298 CTTGGTAAACAAATCTCAAGTGCAGCAAGTTGTCAAGTGCAGCAAGTGCAGCAAGTGC 357  
434 tctcgcgtgctccctg 450  
358 TAAACCCCGACCGCCCG 374

RESULT 15  
BI099174 578 bp mRNA EST 26-JUN-2001  
LOCUS IP1.38.G12.g1\_A002 Immature panicle 1 (IP1) Sorghum bicolor cDNA,  
DEFINITION mRNA sequence.  
ACCESSION BI099174  
VERSION BI099174.1 GI:14570756  
KEYWORDS EST.  
SOURCE sorghum.  
ORGANISM Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 578)  
Klein,R.R., Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M. and Pratt,  
L.H.  
An EST database from Sorghum: developing preanthesis panicles  
unpublished (2001)  
Contact: Cordonnier-Pratt MM  
Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmp@tuega.edu

Sequences have been trimmed to exclude PolyA, vector and regions below phred quality 16. The threshold for high quality sequence is 20. Three-prime sequences, which are obtained with PolyTmX or T7

sequencing primer, are presented as the reverse complement.

Seq primer: T7  
High quality sequence start: 2  
High quality sequence stop: 578  
POLYA-No.

FEATURES  
Source Location/Qualifiers

1..578  
/organism="Sorghum bicolor"  
/cultivar="Brx623"  
/db\_xref="taxon:4558"  
/clone\_11b="Immature panicle 1 (IP1)"  
/note="Organ: Developing preanthesis panicles; Vector: pBluescript II SK(-) from lambda zap II; Site\_1: XhoI; Site\_2: EcoRI. The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."  
BASE COUNT 102 a 164 c 172 g 140 t  
ORIGIN

Query Match 7.18; Score 42.6; DB 11; Length 578;  
Best Local Similarity 47.9%; Pred. No. 0.11;  
Matches 123; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

194 tggcatacaaaagctcgactcgcttcgaagaatcgcccaagaatccaactgaacaaag 253  
109 TGGGTTCGACCCGAGGCGCGGTGTTCACGCGGAGAGATGAGCGCATGAGACGG 168  
254 ctccgagcgaagctgctgttaacgccttcacagatgtgttcgtagagctatgaatcgcg 313  
169 ACTGGGAAGGTTCAAGCCGCCGCCGCCGAGAGAGAGAGATGAGAGAGAGTTCA 228  
314 gcgaaagcctgaagctaacgcgctgttcctccgctgaagcgctgaagcgccgcgcga 373  
229 TGCAGGCTTCACCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 288  
374 ccggccgcaaacccgcgacgagcagagatgagacatcccgcttcctcctgaagcgcttga 433  
289 CTTGGTAAACAAATCTCAAGTGCAGCAAGTTGTCAAGTGCAGCAAGTGCAGCAAGTGC 348  
434 tctcgcgtgctccctg 450  
349 TAAACCCCGACCGCCCG 365

Search completed: December 20, 2001, 19:20:31  
Job time: 6407 sec



ORGANISM: Homo sapiens  
TISSUE TYPE: Placenta  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(38..2375)  
FEATURE:  
NAME/KEY: mat.peptide  
LOCATION: join(101..2375)  
US-08-426-627-3

Query Match 15.2%; Score 29.2; DB 1; Length 3092;  
Best Local Similarity 59.8%; Pred. No. 1.4;  
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 89 AACCTTATTAACGGGGTTTCGAGAAACATGCGCTAGTATCATGTGACAAACATG 148  
DB 626 AACCATATTCCTAATGGGGTGTCACTGTTAATGTGTCGATCAATCATGGAACACAG 685  
QY 149 GACTAGCAAAAGTGTGTC 170  
DB 686 ATTGACAAATGGTGTGTC 707

## RESULT 2

US-08-477-396A-3  
Sequence 3, Application US/08477396A  
Patent No. 5872235  
GENERAL INFORMATION:  
APPLICANT: Chen, Ian Bo  
APPLICANT: Bao, Shideng  
APPLICANT: Liu, Yuan  
TITLE OF INVENTION: A NOVEL TUMOR MARKER AND NOVEL METHOD OF  
TITLE OF INVENTION: ISOLATING SAME  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes  
STREET: Ten Post Office Square  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,396A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US-08/146,488  
FILING DATE: 29-OCT-1993  
APPLICATION NUMBER: US 08/448,388  
FILING DATE: 28-MAY-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/12502  
FILING DATE: 31-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Heine, Holliday C.  
REGISTRATION NUMBER: 34,346  
REFERENCE/DOCKET NUMBER: DFCI-333BX  
TELEPHONE: (617) 542-2290  
TELEFAX: (617) 451-0313  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3126 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO

ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 43..2376  
US-08-477-396A-3

Query Match 15.2%; Score 29.2; DB 2; Length 3126;  
Best Local Similarity 59.8%; Pred. No. 1.4;  
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 89 AACCTTATTAACGGGGTTTCGAGAAACATGCGCTAGTATCATGTGACAAACATG 148  
DB 649 AACCATATTCCTAATGGGGTGTCACTGTTAATGTGTCGATCAATCATGGAACACAG 708  
QY 149 GACTAGCAAAAGTGTGTC 170  
DB 709 ATTGACAAATGGTGTGTC 730

## RESULT 3

US-08-426-627-5  
Sequence 5, Application US/08426627  
Patent No. 5756664  
GENERAL INFORMATION:  
APPLICANT: Amann, Egon  
APPLICANT: Otawara-Hamamoto, Yoko  
APPLICANT: Kikuno, Reiko  
APPLICANT: Takeshita, Sunao  
APPLICANT: Tezuka, Kenichi  
TITLE OF INVENTION: No. 5756664e1 Protein with Bone Formation  
TITLE OF INVENTION: Ability and Process for Its Production.  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/426,627  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/036,841  
FILING DATE: 25-MAR-1993  
APPLICATION NUMBER: JP 4-71501  
FILING DATE: 27-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Hammond, Alan W.  
REGISTRATION NUMBER: 35,178  
REFERENCE/DOCKET NUMBER: 02481-1285-00000  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3253 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL TYPE: osteosarcoma  
FEATURE:  
NAME/KEY: CDS

LOCATION: join(32..2540)  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: join(97..2540)  
US-08-426-627-5

Query Match 15.2%; Score 29.2; DB 1; Length 3253;  
Best Local Similarity 59.8%; Pred. No. 1.4;  
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 89 AACCTTATATAACGGGTTTCACAGAAACATGCGCTAGATCATGATGACAAACATG 148  
DB 620 AACATTATCTTAATGAGGTTCTCATGTATGTGCTCGAATCATCGAGAACG 679

QY 149 GACTAGCAAAAGCTGTGTCC 170  
DB 680 ATTGCACAAATGTGTGTCC 701

## RESULT 4

US-08-943-731-128/c  
Sequence 128, Application US/08943731  
Patent No. 6265157

## GENERAL INFORMATION:

APPLICANT: PROCKOP, DARWIN J.  
APPLICANT: SPOTILA, LORETTA D.  
APPLICANT: DELTAS, CONSTANTINOS D.  
APPLICANT: SEREDA, LARISA  
APPLICANT: LARSON, ANDREA W.  
APPLICANT: PACK, MICHAEL  
APPLICANT: COLIGE, ALAIN  
APPLICANT: EARLY, JAMES  
APPLICANT: KORRKO, JARMO  
APPLICANT: ALA-KORRKO, LEENA, et al.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING  
TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES  
NUMBER OF SEQUENCES: 666  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.  
STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND  
CITY: PHILADELPHIA  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-7086

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/943,731  
FILING DATE: 03-OCT-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/212,322  
FILING DATE: 14-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/803,628  
FILING DATE: 03-DEC-1991

## ATTORNEY/AGENT INFORMATION:

NAME: DOYLE LEARY Ph.D., KATHRYN  
REGISTRATION NUMBER: 36,317  
REFERENCE/DOCKET NUMBER: 9598-27  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-965-1284  
TELEFAX: 215-567-2991  
TELEX: 831-494

## INFORMATION FOR SEQ ID NO: 128:

SEQUENCE CHARACTERISTICS:  
LENGTH: 730 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double

TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-943-731-128

Query Match 14.5%; Score 27.8; DB 4; Length 730;  
Best Local Similarity 65.1%; Pred. No. 2.4;  
Matches 41; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 130 TATCATGTATGACACATGACATAGCAAAAGTGTGTCCCTGACCCAGAAAGATGC 189  
DB 374 TTTCAGAAAGTGCACCTTGTGACCAAGCTCAGCTGCTATCCCTCCCTCGAAGGCTCC 315

QY 190 TTT 192  
DB 314 TCT 312

## RESULT 5

US-08-330-154-1  
Sequence 1, Application US/08330154  
Patent No. 5587307

## GENERAL INFORMATION:

APPLICANT: Alborn Jr., William E  
APPLICANT: Hoskins, Joann  
APPLICANT: Skatrud, Paul L  
APPLICANT: Unal, Serhat  
TITLE OF INVENTION: PEMA GENE OF STAPHYLOCOCCUS EPIDERMIDIS,  
TITLE OF INVENTION: FEMA PROTEIN, AND VECTORS AND MICROORGANISMS COMPRISING  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patent Division/AEH  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: USA  
ZIP: 46285

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/330,154  
FILING DATE: 27-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/208,925  
FILING DATE: 09-MAR-1994  
APPLICATION NUMBER: US/08/057,163  
FILING DATE:

## ATTORNEY/AGENT INFORMATION:

NAME: Hamillon, Amy E  
REGISTRATION NUMBER: 33,894  
REFERENCE/DOCKET NUMBER: X-8894  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-3169  
TELEFAX: 317-276-1294

## INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1257 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:

NAME/KEY: CDS  
LOCATION: 1..1257  
US-08-330-154-1

Query Match 14.5%; Score 27.8; DB 1; Length 1257;  
Best Local Similarity 62.0%; Pred. No. 3;  
Matches 44; Conservative 0; Mismatches 27; Indels 0; Gaps 0;



```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 911:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7650 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...7650
; US-09-221-017B-911

Query Match
Best Local Similarity 14.3%; Score 27.4; DB 4; Length 7650;
Matches 58; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 39 AACTTCGATTCGCTATTTTCATACCTTGGGGAATTAATGTGAACCCCTTATA 98
DB 1504 AACTGCCCAAACTTCGATTCGCTACCTCCGCTTTTAAACAGGATCGGCCCTTAA 1445
QY 99 AAACGGGGTTTTCGCAAAACATGCGCTGATTCATTTGATGACAAACAT 147
DB 1444 AGACGGGTCTCTTCGGCGAAAGAGCTGAGCATCTCGATGCGACGCTT 1436

RESULT 9
US-09-426-290-1
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Berghild Ran Olafsdottir
; APPLICANT: Jeffrey Gulcher
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/09/426,290
; CURRENT FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 168575
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21181)...(21403)
; NAME/KEY: CDS
; LOCATION: (95252)...(95430)
; NAME/KEY: CDS
; LOCATION: (101753)...(101966)
; NAME/KEY: CDS
; LOCATION: (110324)...(110439)
; NAME/KEY: CDS
; LOCATION: (124058)...(124278)
; NAME/KEY: CDS
; LOCATION: (127009)...(127130)
```

```

; NAME/KEY: CDS
; LOCATION: (128910)...(129139)
; US-09-426-290-1

Query Match
Best Local Similarity 14.2%; Score 27.2; DB 4; Length 168575;
Matches 56; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 56 TATTTCAATACCTTCGGGGAATAGATGGAACCCCTATATAACGGCGTTTCGCA 115
DB 109602 TATTCATATGATGATGAGAAATTAATGTTATTAAGCCTTAATAATTCATTTTATTA 109661
QY 116 GAAACATGCGCTAGTATTCATTTGATGACAAACATGACTAAGCAAA 159
DB 109662 AAGTATATATACATGATTTTATTAATAAGCATACACACACACACA 109705

RESULT 10
5223394-8
; Patent No. 5223394
; APPLICANT: WALLNER, BARBARA
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULE COMPRISING
; LYMPHOCYTE FUNCTION-ASSOCIATED ANTIGEN 3 PHOSPHATIDYLINOSITOL
; LINKAGE SIGNAL SEQUENCE
; NUMBER OF SEQUENCES: 12
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/335,688
; FILING DATE: 10-APR-1989
; SEQ ID NO: 8
; LENGTH: 1009
5223394-8

Query Match
Best Local Similarity 14.1%; Score 27; DB 6; Length 1009;
Matches 48; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 74 GGAATATAGATGTAAGAAACCTTATATAAAGCGGCTTTCGCAAGAACATGCGCTAGTATC 133
DB 777 GGAAGTGTCTGTAAAGCGGTTTACAAAGACACAGATATGACATTTATACCATACATTAGC 836
QY 134 ATTGATGACACATGATGATGAC 156
DB 837 AGTATTAACAACATGATGTGTC 859

RESULT 11
US-08-204-288-3/c
; Sequence 3, Application US/08204288
; Patent No. 5959178
; GENERAL INFORMATION:
; APPLICANT: VAN DOORSELAERE, Jan
; APPLICANT: FRITIG, Bernard J.M.
; APPLICANT: INZE, Dirk G.
; APPLICANT: JOUANIN, Lise
; APPLICANT: KNIGHT, Mary E.
; APPLICANT: VAN MONTAGU, Marc
; APPLICANT: LEGRAND, Michel
; TITLE OF INVENTION: MODIFICATION OF LIGNIN SYNTHESIS IN
; PLANTS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DABY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20005-3518
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/204,288  
FILING DATE: 10-MAR-1994  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9119279.9  
FILING DATE: 10-SEP-1991  
PRIOR APPLICATION DATA: PCT/GB92/01460  
APPLICATION NUMBER: PCT/GB92/01460  
FILING DATE: 09-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, Paul N.  
REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: 206860/SPE36543/UST  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1244 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-204-288-3

Query Match 14.1%; Score 27; DB 2; Length 1244;  
Best Local Similarity 62.7%; Pred. No. 5.6;  
Matches 42; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

OY 57 ATTTCATACCTTGGGGAATAGATGTGAAAACCTTTATAAACGGGGTTTTCGAC 116  
DB 564 ATCTCAATACATGTGGTAATCAAGTTATACCTTAATAGAGGATGTTTGAGAC 505  
OY 117 AACATG 123  
DB 504 AATCATG 498

RESULT 12  
US-08-188-582-12/C  
Sequence 12, Application US/08188582  
Patent No. 5534410  
GENERAL INFORMATION:  
APPLICANT: Tjian, Robert  
APPLICANT: Comai, Lucio  
APPLICANT: Dynlacht, Brian D.  
APPLICANT: Hoey, Timothy  
APPLICANT: Ruppert, Siegfried  
APPLICANT: Tanese, Naoko  
APPLICANT: Wang, Edith  
APPLICANT: Weinzierl, Robert O.J.  
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,  
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/188,582  
FILING DATE: 28-JAN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman, Richard A

REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-57650-2/ALT/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3182 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 972..3002  
US-08-188-582-12

Query Match 14.1%; Score 27; DB 1; Length 3182;  
Best Local Similarity 66.1%; Pred. No. 8.2;  
Matches 39; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

OY 2 CTGGCGCGGCGGCATGAGTGGCTTGACACATATCTTGCTGATTCGTATTT 60  
DB 1238 CTGGCGTGAAGCATAGAGTGGCTGCACATCTTTAGCTTCAAGCGGTAGCATGT 1180

RESULT 13  
US-08-646-715-12/C  
Sequence 12, Application US/08646715  
Patent No. 5637686  
GENERAL INFORMATION:  
APPLICANT: Tjian, Robert  
APPLICANT: Comai, Lucio  
APPLICANT: Dynlacht, Brian D.  
APPLICANT: Hoey, Timothy  
APPLICANT: Ruppert, Siegfried  
APPLICANT: Tanese, Naoko  
APPLICANT: Wang, Edith  
APPLICANT: Weinzierl, Robert O.J.  
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,  
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/646,715  
FILING DATE: 09-MAY-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/188,582  
FILING DATE: 28-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman, Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-57650-2/ALT/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3182 base pairs



TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 972..3002  
US-08-646-715-12

Query Match 14.1%; Score 27; DB 1; Length 3182;

Best Local Similarity 66.1%; Pred. No. 8.2;  
Matches 39; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 2 CTGGCGCGCGCGCATGATGCTTGGACAAGCATATCTGTCTGATTGCTATTT 60  
DB 1238 CTGGCGCGTGAAGCCATTAAGTGGCTGCGACATCTTACTCTCAAGCGCTAGCAATGT 1180

RESULT 14

US-08-306-546C-1  
Sequence 1, Application US/08306546C  
Patent No. 5605797

GENERAL INFORMATION:

APPLICANT: Friderich, Karen

APPLICANT: Jones, Margaret

APPLICANT: Chen, Hong

APPLICANT: Cavanagh, Kevin

TITLE OF INVENTION: Bovine Beta-Mannosidase Gene and Methods

TITLE OF INVENTION: of Use

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESSES:

ADDRESS: Harness, Dickey & Pierce, P.L.C.

STREET: P.O. Box 828

CITY: Bloomfield Hills

STATE: Michigan

COUNTRY: USA

ZIP: 48303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/306,546C

FILING DATE: September 15, 1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Smith, Deann F.

REGISTRATION NUMBER: 36,683

REFERENCE/DOCKET NUMBER: 6550-00003

TELECOMMUNICATION INFORMATION:

TELEPHONE: (810)641-1600

TELEFAX: (810)641-0270

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3852 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 75..2711

US-08-546C-1

Query Match 14.1%; Score 27; DB 1; Length 3852;

Best Local Similarity 56.0%; Pred. No. 8.8;  
Matches 51; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 75 GAAATAGATGCGAAGACCTTATAAAGCGGCTTTGGCAGAAACATCGCTAGATATCA 134

DB 3567 GATTTCATGTGATGAGCTGGACGACCAAGTGTATGCTTAACTTGATGATGACCA 3626

QY 135 TTGATGACACATGACCTAAGCAAAAGTGT 165  
DB 3627 TTTTAAAAACACTGATATATATATTGCT 3657

RESULT 15

US-08-530-524A-1  
Sequence 1, Application US/08530524A  
Patent No. 5837836

GENERAL INFORMATION:

APPLICANT: Friderich, Karen

APPLICANT: Jones, Margaret

APPLICANT: Chen, Hong

APPLICANT: Cavanagh, Kevin

TITLE OF INVENTION: Bovine Beta-Mannosidase Gene and Methods

TITLE OF INVENTION: of Use

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESSES:

ADDRESS: Harness, Dickey & Pierce, P.L.C.

STREET: P.O. Box 828

CITY: Bloomfield Hills

STATE: Michigan

COUNTRY: USA

ZIP: 48303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/530,524A

FILING DATE: September 19, 1995

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Smith, Deann F.

REGISTRATION NUMBER: 36,683

REFERENCE/DOCKET NUMBER: 6550-00003DVA

TELECOMMUNICATION INFORMATION:

TELEPHONE: (810)641-1600

TELEFAX: (810)641-0270

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3852 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 75..2711

US-08-530-524A-1

Query Match 14.1%; Score 27; DB 2; Length 3852;

Best Local Similarity 56.0%; Pred. No. 8.8;  
Matches 51; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 75 GAAATAGATGCGAAGACCTTATAAAGCGGCTTTGGCAGAAACATCGCTAGATATCA 134

DB 3567 GATTTCATGTGATGAGCTGGACGACCAAGTGTATGCTTAACTTGATGATGACCA 3626

QY 135 TTGATGACACATGACCTAAGCAAAAGTGT 165

DB 3627 TTTTAAAAACACTGATATATATATTGCT 3657

Search completed: December 25, 2002, 12:58:44  
Job time: 144.664 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 25, 2002, 11:24:48 ; Search time 46.0561 Seconds  
(without alignments)  
1693.523 Million cell updates/sec

Title: US-09-816-391A-1\_COPY\_1\_192

Perfect score: 192  
Sequence: 1 gctggcgccgagcgccatga.....tgaccacgaagaagtgttt 192

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 363474 seqs, 203117208 residues

Total number of hits satisfying chosen parameters: 726948

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_NA.\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	192	100.0	600	10	US-09-816-391A-1
2	30.2	15.7	515	10	US-09-864-761-16206
3	29.2	15.2	3202	10	US-09-925-301-471
4	29.2	15.0	6465	10	US-09-954-456-193
5	28.8	15.0	413	10	US-09-867-701-2315
6	28.8	15.0	419	10	US-09-783-590-663
7	28.4	14.8	345	10	US-09-974-300-3770
8	28.4	14.8	75899	10	US-09-854-883-243
9	27.6	14.4	727	10	US-09-770-149-122
10	27.4	14.3	484	10	US-09-747-155-261
11	27.4	14.3	2000	9	US-09-938-842A-448
12	27.4	14.3	202001	10	US-09-734-674-3
13	27	14.1	389	10	US-09-867-701-2306
14	27	14.1	1577	10	US-09-864-761-30755
15	27	14.1	1870	9	US-09-938-842A-3414
16	27	14.1	1955	10	US-09-864-761-14202
17	27	14.1	2725	10	US-09-864-824A-243
18	27	14.1	2725	10	US-09-880-107-2181
19	27	14.1	14070	9	US-09-870-759-127

20	27	14.1	14635	10	US-09-880-107-2287	Sequence 2287, Ap
21	26.8	14.0	1710	10	US-09-879-957-189	Sequence 189, Ap
22	26.8	14.0	1782	10	US-09-974-300-2801	Sequence 2801, Ap
23	26.6	13.9	182	10	US-09-783-590-4530	Sequence 4530, Ap
24	26.4	13.7	2324	10	US-09-771-161A-74	Sequence 74, Appl
25	26.4	13.7	2544	10	US-09-771-161A-74	Sequence 74, Appl
26	26.4	13.7	32038	10	US-09-764-878-292	Sequence 8723, Ap
27	26.2	13.6	176	10	US-09-867-701-8273	Sequence 8273, Ap
28	26.2	13.6	855	10	US-09-974-300-998	Sequence 998, Ap
29	26.2	13.6	2000	9	US-09-938-842A-22759	Sequence 2759, Ap
30	26.2	13.6	2005	10	US-09-939-980-102	Sequence 102, Ap
31	26.2	13.6	4527	10	US-09-901-940-3	Sequence 102, Ap
32	26.2	13.6	4594	10	US-09-070-922A-220	Sequence 230, Appl
33	26.2	13.6	63000	9	US-09-780-172-18	Sequence 18, Appl
34	26.2	13.5	402	9	US-09-933-797-133	Sequence 133, Appl
35	26	13.5	490	10	US-09-604-287A-95	Sequence 95, Appl
36	26	13.5	490	10	US-09-339-338-95	Sequence 95, Appl
37	26	13.5	490	12	US-10-007-805-95	Sequence 95, Appl
38	26	13.5	546	10	US-09-777-564-1613	Sequence 1613, Ap
39	26	13.5	814	9	US-10-001-857-5	Sequence 5, Appl
40	26	13.5	900	10	US-09-974-300-2838	Sequence 2838, Ap
41	26	13.5	1365	12	US-10-073-256-49	Sequence 49, Appl
42	26	13.5	302250	10	US-09-962-832-154	Sequence 154, Appl
43	25.8	13.4	201	10	US-09-783-590-5925	Sequence 5925, Ap
44	25.8	13.4	391	10	US-09-864-761-4151	Sequence 4151, Ap
45	25.8	13.4	397	10	US-09-960-352-2723	Sequence 2723, Ap

## ALIGNMENTS

RESULT 1  
US-09-816-391A-1  
Sequence 1, Application US/09816391A  
Patent No. US20020054865A1  
GENERAL INFORMATION:  
APPLICANT: FUJIMORI, Minoru  
APPLICANT: TANIGUCHI, Shunichiro  
APPLICANT: AMANO, Jun  
APPLICANT: YAZAWA, Kazuyuki  
APPLICANT: KANO, Yasunobu  
APPLICANT: NAKAMURA, Toshiyuki  
APPLICANT: SAKAKI, Takayuki  
TITLE OF INVENTION: Anaerobic bacterium as a drug for cancer gene therapy  
FILE REFERENCE: 2001-WM/01736  
CURRENT APPLICATION NUMBER: US/09/816,391A  
CURRENT FILING DATE: 2001-03-26  
PRIORITY FILING DATE: 2000-09-21  
PRIORITY FILING DATE: 2000-09-21  
NUMBER OF SEQ ID NOS: 3  
SEQ ID NO 1  
LENGTH: 600  
TYPE: DNA  
ORGANISM: Bifidobacterium longum  
NAME/KEY: CDS  
LOCATION: (193)..(471)  
US-09-816-391A-1

Query Match  
Best Local Similarity 100.0%; Score 192; DB 10; Length 600;  
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCTGGCGCGCGCCATGAGTGGCTTGACACATTAATCTGTCGATCGTCATTT 60  
DB 1 GCTGGCGCGCGCGCCATGAGTGGCTTGACACATTAATCTGTCGATCGTCATTT 60  
QY 61 TCAATACCTTCGGGGAATAGATGTGAACCCCTTATTAACCGGGGTTTTCGAGAAC 120  
DB 61 TCAATACCTTCGGGGAATAGATGTGAACCCCTTATTAACCGGGGTTTTCGAGAAC 120  
QY 121 ATGGCTAGATCATTTATGACACATGACACTAAGCAAAAGCTTCCCTGACCCAA 180  
DB 121 ATGGCTAGATCATTTATGACACATGACACTAAGCAAAAGCTTCCCTGACCCAA 180

Dd	121	ATGGCGTAGTATCTTGTGATGACACATGGACTAAGCAAAAAGTGGTTCCTCCCTGACCCA	180
Qy	181	GAAGGATGCTTT	192
Dd	181	GAAGGATGCTTT	192

RESULT 2  
US-09-86

; Sequence 16206, Application US/09864761  
; Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Chen, Shaolin G.  
 APPLICANT: Rank, David R.  
 APPLICANT: Hanzel, David K.  
 APPLICANT: Chen, Wensheng  
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
 FILE REFERENCE: Aecmcca-X-1  
 CURRENT APPLICATION NUMBER: US/09/864,761

?  
 ? CURRENT FILING DATE: 2001-05-23  
 ? PRIOR APPLICATION NUMBER: US 60/180,312  
 ? PRIOR FILING DATE: 2000-02-04  
 ? PRIOR APPLICATION NUMBER: US 60/207,456  
 ? PRIOR FILING DATE: 2000-05-26  
 ? PRIOR APPLICATION NUMBER: US 09/632,366  
 ? PRIOR FILING DATE: 2000-08-03  
 ? PRIOR APPLICATION NUMBER: GB 24263.6  
 ? PRIOR FILING DATE: 2000-10-04  
 ?  
 ?

1 PRIOR APPLICATION NUMBER: US 60/236,359  
 2  
 3 PRIOR FILING DATE: 2000-09-27  
 4  
 5 PRIOR APPLICATION NUMBER: PCT/US01/00666  
 6  
 7 PRIOR FILING DATE: 2001-01-30  
 8  
 9 PRIOR APPLICATION NUMBER: PCT/US01/00667  
 10  
 11 PRIOR FILING DATE: 2001-01-30  
 12  
 13 PRIOR APPLICATION NUMBER: PCT/US01/00664  
 14  
 15 PRIOR FILING DATE: 2001-01-30  
 16  
 17 PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30

1 PRIOR APPLICATION NUMBER: PCT/US01/00661  
 2  
 3 PRIOR FILING DATE: 2001-01-30  
 4  
 5 PRIOR APPLICATION NUMBER: PCT/US01/00670  
 6  
 7 PRIOR FILING DATE: 2001-01-30  
 8  
 9 PRIOR APPLICATION NUMBER: US 60/234,687  
 10  
 11 PRIOR FILING DATE: 2000-09-21  
 12  
 13 PRIOR APPLICATION NUMBER: US 09/608,408  
 14  
 15 PRIOR FILING DATE: 2000-06-30  
 16  
 17 PRIOR APPLICATION NUMBER: US 09/774,702

```

? PRIOR APPLICATION NUMBER: US 05/114,203
? PRIOR FILING DATE: 2001-01-29
? NUMBER OF SEQ ID NOS: 49117
? SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
? SEQ ID NO 16206
?   LENGTH: 515
?   TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:

```

```

; OTHER INFORMATION: MAP TO AC000062.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
US-09-864-761-16206

Query Match          15.7%   Score 30.2; DB 10; Length 515;
Best Local Similarity 49.7%   Pred. No. 0.57;
Matches 77; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

OY 13 GCCCATGAGAGTGGCTTGACAGACATATCTGTGTCGATTCGTCCTATTTCATACCTTCG 72

```

Db 253 GCCCTTAAATGACACAGTAGATATTATTCCTTAGTGAGGACAGTTTAAAAATCCTTA 312

Qy 73 GGGAAATGATGATGAAAAACCTTATAAAACGCGGTTTTGCAGAAAACGCTAGTAT 133

Db 313 TAGAAGTGATGATTAGACATTCACAATTAATGTTTGTATCATTCAGCTGTGTTTT 372

Qy 133 CATGATGACACATGACGTAAAGCAAGTCTTG 167  
||| | | ||| | | | | |  
Db 373 GTTTACGCGCTTTAAACTTTTAAAGACTTCG 407

### RESULT 3

US-09-9425-501-471  
; Sequence 471, Application US/09925301  
; Patent No. US20020052308A1  
: GENERAL INFORMATION:

APPLICANT: Rosen et al

```

: TITLE OF INVENTION: Nucleic acids, proteins and antibodies
: FILE REFERENCE: PA106
: CURRENT APPLICATION NUMBER: US/09/925, 301
: CURRENT FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: PCT/US00/05882
: PRIOR FILING DATE: 2000-03-08
: PRIOR APPLICATION NUMBER: 60/124, 270
: PRIOR FILING DATE: 1999-03-12
: NUMBER OF SEQ ID NOS: 1694
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 471

```

```

?      LENGTH: 3202
?      TYPE: DNA
?      ORGANISM: Homo sapiens
?      FEATURE:
?      NAME/KEY: misc_feature
?      LOCATION: (4)
?      OTHER INFORMATION: n equals a,t,g, or c
?      NAME/KEY: misc_feature
?      LOCATION: (3160)
?
```

```

; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (3180)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-471

Query Match          15.2%   Score 29.
Best local Similarity 59.8%   Pred. No.
Matches 49: Conservative 0: Mismatch

```

[illegible]

## RESULT 4

US-09-954-456-193  
; Sequence 193, Application US/09954456  
; Patent No. US20020115057A1  
; Patent No. US20020115057A1

GENERAL INFORMATION  
APPLICANT: YOUNG

/ TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using  
 /  
 / TITLE OF INVENTION: Sels  
 / FILE REFERENCE: 688290-76  
 /  
 / CURRENT APPLICATION NUMBER: US/09/954,456  
 /  
 / CURRENT FILING DATE: 2001-09-18  
 /  
 / PRIOR APPLICATION NUMBER: US/60/233,617  
 /  
 / PRIOR FILING DATE: 2000-09-18  
 /  
 / PRIOR APPLICATION NUMBER: US/60/234,052  
 /  
 / PRIOR FILING DATE: 2000-09-20  
 /  
 / PRIOR APPLICATION NUMBER: US/60/234,923  
 /



```
;; LOCATION: (315)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (332)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (335)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (338)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (340)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (348)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (357)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (375)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (380)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (381)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (390)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (394)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (397)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (413)
;; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-663

Query Match 15.0%; Score 28.8; DB 10; Length 419;
Best Local Similarity 50.4%; Pred. No. 1.6;
Matches 60; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 42 TTGCTGATTCCTATTTTCAATACCTCGGGAAATAGATGTGAAAACCTTATATAA 101
DB 175 TTTCCTTTGCTCTTAATCTCAACCTTTGGAGACCAAGGTGGGATGGCTTGAAC 234
QY 102 CGCGGGTTTTCGAGAAACATGCGCTAGTATCATTTGAGACAAATGACTAAGCAAA 160
DB 235 CAGGAGTTTNTGANCACCCCTGGGNNANCATAGTGAGNACCACATGTTTACAAANA 293

RESULT 7
US-09-974-300-3770/c
; Sequence 3770, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Id Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085, 500-US
; CURRENT APPLICATION NUMBER: US/09/974, 300
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 09/680, 598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279, 526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
```

```
;; SOFTWARE: PastSeq for Windows Version 4.0
;; SEQ ID NO 3770
;; LENGTH: 345
;; TYPE: DNA
;; ORGANISM: Bacillus licheniformis
US-09-974-300-3770

Query Match 14.8%; Score 28.4; DB 10; Length 345;
Best Local Similarity 49.3%; Pred. No. 2;
Matches 74; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 17 ATGAAGTGGCTTGACAAACATATCTGTGATTCGTCATTTTCAATACCTTCGGGGA 76
DB 258 ATGATTTTATTTGGACGATGATCTTATCCGTCTTTTGTCTATTTTCCGCA 199
QY 77 AATGATGTGAACCCCTTATPAACCGCGGTTTTCGAGAACATCGCTATTCATT 136
DB 198 AAGAGATTTTTCACAAATTTACAAAGTTGTTCAATTCATATTTGAACAGCTTTT 139
QY 137 GATGACATGAGCTAAGCAAAAGTCTT 166
DB 138 TGTCTTAATAGATTAAGACGCTTTT 109

RESULT 8
US-09-854-883-243/c
; Sequence 243, Application US/09854883
; Patent No. US20020053479A1
; GENERAL INFORMATION:
; APPLICANT: Lex M. Cowser
; APPLICANT: Jacqueline Wyatt
; APPLICANT: Susan M. Freier
; APPLICANT: Brett P. Monia
; APPLICANT: Madeline M. Butler
; APPLICANT: Robert McKay
; TITLE OF INVENTION: ANTISENSE MODULATION OF PTP1B EXPRESSION
; FILE REFERENCE: ISPH-0576
; CURRENT APPLICATION NUMBER: US/09/854, 883
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 09/629, 644
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 09/487, 368
; PRIOR FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 389
; SEQ ID NO 243
; LENGTH: 75899
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-854-883-243

Query Match 14.8%; Score 28.4; DB 10; Length 75899;
Best Local Similarity 58.1%; Pred. No. 26;
Matches 50; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 57 ATTTCATACCTTCGGGAATAGATGTGAACCCCTTATPAACGGGGTTTTCGAG 116
DB 16542 ATTTTCACTTCCTCCGAAATCAGAAAACAAAGCATATCAACTGGGCTTCTGCA 16483
QY 117 AAACATGCGCTAGTATCATTTGATGAC 142
DB 16482 GAGTTTGGCTTCTATCCAGATGCC 16457

RESULT 9
US-09-770-149-122/c
; Sequence 122, Application US/09770149
; Patent No. US20020059663A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
```



;  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647

; PRIOR APPLICATION NUMBER: US 60/300,111  
 ; PRIOR FILING DATE: 2001-06-22  
 ; NUMBER OF SEQ ID NOS: 5379  
 ; SEQ ID NO 3414  
 ; LENGTH: 1870  
 ; TYPE: DNA  
 ; ORGANISM: Arabidopsis thaliana  
 US-09-938-842A-3414

Query Match 14.1%; Score 27; DB 9; Length 1870;  
 Best Local Similarity 53.3%; Pred. No. 14;  
 Matches 57; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
 QY 29 GACACGATATCTGCTGATCGTCTATTTTCATATCCTTCGGGGAATAGATGTGAA 88  
 DB 1599 GAAAAAGTGTATTTATTTAGGCCATATAATACCATATTTGAAATAGATATTTT 1540  
 QY 89 AACCTTATATAACGGGGGTTTCGCAGAAACATCGCTAGATCAT 135  
 DB 1539 AGTACATATTTTGGGGGTTCATGTAAACAAATAATATCAT 1493

Search completed: December 25, 2002, 12:34:27  
 Job time : 178.056 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 25, 2002, 11:24:49 ; Search time 29.3364 Seconds  
(without alignments)  
1348.538 Million cell updates/sec

Title: US-09-816-391a-1\_COPY\_472\_600

Sequence: 129  
1 tgcaccttcgtcgtcgtacga.....tattccgactagcttcagcg 129

Scoring table: IDENTITY\_NIC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/1/lna/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/lna/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/lna/5A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/lna/5B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/lna/5A\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/lna/5B\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	28.4	22.0	2 US-08-602-359A-31	Sequence 31, Appl
2	26.2	20.3	3 US-08-672-850-7	Sequence 7, Appl
3	26.2	20.3	3 US-08-672-850-11	Sequence 11, Appl
4	26.2	20.3	3 US-08-672-850-11	Sequence 10, Appl
5	26.2	20.3	3 US-08-672-850-11	Sequence 11, Appl
6	25.8	20.0	4 US-08-804-227C-1	Sequence 1, Appl
7	25.8	20.0	4 US-08-198-092-1	Sequence 1, Appl
8	25.8	20.0	2 US-08-674-887A-5	Sequence 5, Appl
9	25.8	20.0	2 US-08-951-844-5	Sequence 5, Appl
10	25.8	20.0	2 US-09-412-347-5	Sequence 5, Appl
11	25.4	19.7	3 US-08-461-244-1	Sequence 1, Appl
12	25.2	19.5	4 US-09-221-017B-174	Sequence 174, Appl
13	25.2	19.5	4 US-08-247-901C-1	Sequence 1, Appl
14	25.2	19.5	4 US-09-075-904-1	Sequence 1, Appl
15	25.2	19.5	4 US-09-426-436-1	Sequence 1, Appl
16	25.2	19.5	4 US-08-705-557-1	Sequence 1, Appl
17	24.8	19.2	4 US-08-483-101-1	Sequence 1, Appl
18	24.8	19.2	4 US-08-173-510B-100	Sequence 100, Appl
19	24.8	19.2	4 US-08-458-218-98	Sequence 98, Appl
20	24.8	19.2	4 US-08-450-497-100	Sequence 100, Appl
21	24.2	18.8	4 US-09-134-001C-2784	Sequence 2784, Appl
22	24.2	18.8	4 US-08-991-677-3	Sequence 3, Appl
23	23.8	18.4	4 US-09-442-100-1	Sequence 1, Appl
24	23.8	18.4	4 US-08-264-578-7	Sequence 7, Appl
25	23.8	18.4	4 US-08-231-193A-41	Sequence 41, Appl
26	23.8	18.4	4 US-08-486-273A-41	Sequence 41, Appl
27	23.8	18.4	4 US-08-480-474-41	Sequence 41, Appl

28	23.8	18.4	2326	4 US-08-940-035A-41	Sequence 41, Appl
29	23.8	18.4	2326	4 US-08-935-105A-41	Sequence 41, Appl
30	23.8	18.4	2326	4 US-09-648-797-41	Sequence 41, Appl
31	23.8	18.4	3243	2 US-08-231-193A-44	Sequence 44, Appl
32	23.8	18.4	3243	2 US-08-486-273A-44	Sequence 44, Appl
33	23.8	18.4	3243	3 US-08-480-474-44	Sequence 44, Appl
34	23.8	18.4	3243	3 US-08-940-086A-44	Sequence 44, Appl
35	23.8	18.4	3243	4 US-08-940-035A-44	Sequence 44, Appl
36	23.8	18.4	3243	4 US-08-935-105A-44	Sequence 44, Appl
37	23.8	18.4	3243	4 US-09-648-797-44	Sequence 44, Appl
38	23.8	18.4	3698	2 US-08-231-193A-43	Sequence 43, Appl
39	23.8	18.4	3698	2 US-08-486-273A-43	Sequence 43, Appl
40	23.8	18.4	3698	3 US-08-480-474-43	Sequence 43, Appl
41	23.8	18.4	3698	4 US-08-940-086A-43	Sequence 43, Appl
42	23.8	18.4	3698	4 US-08-935-105A-43	Sequence 43, Appl
43	23.8	18.4	3698	4 US-09-648-797-43	Sequence 43, Appl
44	23.8	18.4	4002	2 US-08-231-193A-53	Sequence 53, Appl

#### ALIGNMENTS

RESULT 1  
US-08-602-359A-31  
Sequence 31, Application US/08602359A  
Patent No. 5942430  
GENERAL INFORMATION:  
APPLICANT: ROBERTSON, Daniel E.  
APPLICANT: MURPHY, Dennis  
APPLICANT: REID, John  
APPLICANT: MAFFIA, Anthony  
APPLICANT: LINK, Steven  
APPLICANT: SWANSON, Ronald V.  
APPLICANT: WARREN, Patrick V.  
APPLICANT: KOSMOTKA, Anna  
TITLE OF INVENTION: ESTERASES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & RICHARDSON P.C.  
STREET: 4225 EXECUTIVE SQUARE, STE 1400  
CITY: LA JOLLA  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,359A  
FILING DATE: February 16, 1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: HAILE, LISA A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09010/010001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-678-5070  
TELEFAX: 619-678-5099  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 936 NUCLEOTIDES  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: GENOMIC DNA  
US-08-602-359A-31

```
Query Match 22.0%; Score 28.4; DB 2; Length 936;
Best Local Similarity 56.4%; Pred. No. 0.46;
Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

OY 2 GACCTCTCTCTAGCATTACTGAGCATTAGCAGCAAGACCCGACGCGGAT 61
DB 718 GACCTTGAAGACCTACTCTCTGCGTCATCATACCGCAATACGACCGCTGAGAGAT 777

OY 62 GGTGGGGGCTTTTGTGTCGTGTGACGTG 95
DB 778 GAAGAGAGAGTTTTCGGGACAGATCTGAGAGAG 811

RESULT 2
US-08-672-850-7/c
; Sequence 7, Application US/08672850
; Patent No. 6140117
; GENERAL INFORMATION:
; APPLICANT: Milbrandt, Jeffrey
; APPLICANT: Arai, Toshiyuki
; TITLE OF INVENTION: NINJURIN
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/672,850
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-63610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-672-850-7

Query Match 20.3%; Score 26.2; DB 3; Length 907;
Best Local Similarity 58.2%; Pred. No. 2.9;
Matches 46; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

OY 11 CTCGTGCGATTCTCTGCGATTACTGAGCAAGACCCGACGATGTCGGGAT 70
DB 434 CCCGAAGGCTGTAATAAATGATGACACAGTGAAGACCAAGATGTGCTGC 375

OY 71 CTTTGTGTGTGCTGT 89
DB 374 GTGTGTGAGCTGTGAGT 356

RESULT 3
US-08-672-850-11/c
; Sequence 11, Application US/08672850
; Patent No. 6140117
; GENERAL INFORMATION:
; APPLICANT: Milbrandt, Jeffrey
```

```
APPLICANT: Arai, Toshiyuki
; TITLE OF INVENTION: NINJURIN
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/672,850
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-63610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 398-3249
; TELEFAX: (415) 781-1989
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-672-850-11

Query Match 20.3%; Score 26.2; DB 3; Length 907;
Best Local Similarity 58.2%; Pred. No. 2.9;
Matches 46; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

OY 11 CTCGTGCGATTCTCTGCGATTACTGAGCAAGACCCGACGATGTCGGGAT 70
DB 434 CCCGAAGGCTGTAATAAATGATGACACAGTGAAGACCAAGATGTGCTGC 375

OY 71 CTTTGTGTGTGCTGT 89
DB 374 GTGTGTGAGCTGTGAGT 356

RESULT 4
US-08-672-850-10/c
; Sequence 10, Application US/08672850
; Patent No. 6140117
; GENERAL INFORMATION:
; APPLICANT: Milbrandt, Jeffrey
; APPLICANT: Arai, Toshiyuki
; TITLE OF INVENTION: NINJURIN
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/672,850
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
```

ATTORNEY/AGENT INFORMATION:  
NAME: SILVA, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-63610  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1001 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-08-672-850-10

Query Match  
Best Local Similarity 20.3%; Score 26.2; DB 3; Length 1001;  
Matches 46; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 11 CTCGACGCAATTAATCTCGACATTAAGACGACCAAGACCCGACGAGATGCTGGGT 70  
DB 528 CCGGAGGCTGTAATCAAAATGATGACACAGTGAAGAAGACCAAGATGCTGCTCG 469  
QY 71 CTTTGTGTGTGCTGCTGT 89  
DB 468 GTTGTGACCTGCTGTGAGT 450

RESULT 5  
US-08-804-227C-1/c  
Sequence 1, Application US/08804227C  
Patent No. 5876991  
GENERAL INFORMATION:  
APPLICANT: Dehoff, Bradley S.  
APPLICANT: Kustoss, Stuart A.  
APPLICANT: Rostock, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804.227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 43280 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
NAME/KEY: CDS  
LOCATION: 816..14234  
FEATURE:

NAME/KEY: CDS  
LOCATION: 14351..19945  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 20010..31199  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31232..36067  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 36249..41774  
US-08-804-227C-1

Query Match  
Best Local Similarity 20.2%; Score 26; DB 2; Length 43280;  
Matches 50; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 38 GACGACAAAGACCCGACGAGATGCTGCTTTTGTGTGCTGACGCTGT 97  
DB 13262 GACCGGAAACACCGACAGCGGCTGCTCCGGTCACTTGTGAGGTTCAGCGCGCTC 13203  
QY 98 GTCCAAACCTAATTAATTCGCGACTACTTCAG 127  
DB 13202 CGCTTCGGCCGACATGACGCTACCATCCG 13173

RESULT 6  
US-09-198-092-1/c  
Sequence 1, Application US/09198092A  
Patent No. 6214575  
GENERAL INFORMATION:  
APPLICANT: Yano, Masamitsu  
APPLICANT: Omura, Mitsuo  
APPLICANT: Ikoma, Yoshihori  
APPLICANT: Komatsu, Akira  
TITLE OF INVENTION: BETA-CAROTENE HYDROXYLASE GENE  
FILE REFERENCE: 07898/033001  
CURRENT FILING DATE: 1998-11-23  
EARLIER APPLICATION NUMBER: US/09/198.092A  
EARLIER FILING DATE: 1997-12-02  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 1  
LENGTH: 1158  
TYPE: DNA  
ORGANISM: Citrus unshiu  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (87)..(1019)  
US-09-198-092-1

Query Match  
Best Local Similarity 20.0%; Score 25.8; DB 4; Length 1158;  
Matches 39; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 37 TGACGACAAAGACCCGACGAGATGCTGCGGCTTTTGTGTGCTGACGCTGT 96  
DB 111 TGGGGCCAAATAGTCGACCGCCATGCTGGGGTTTGTGTTTACAGAGTTTGGGTTT 52  
QY 97 T 97  
DB 51 T 51

RESULT 7  
US-08-674-887A-5/c  
Sequence 5, Application US/08674887A  
Patent No. 5939300  
GENERAL INFORMATION:  
APPLICANT: Robertson, Dan E.  
APPLICANT: Sanjay, Indrajit  
APPLICANT: Adhikari, Robert S.

TITLE OF INVENTION: CATALASES  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/674,887A  
FILING DATE: 03-JUL-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Hallie, Ph.D., Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09015/002001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2262 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 1...2259  
US-08-674-887A-5

Query Match 20.0%; Score 25.8; DB 2; Length 2262;  
Best Local Similarity 60.9%; Pred. No. 5.9; Mismatches 27; Indels 0; Gaps 0;  
Matches 42; Conservative 0;

QY 13 CGTAGCATTTACTTCGACGATTACGACACCAAGACCCGACGAGATGTCGGGGTCT 72  
DB 916 CATTCGCAATTACCGGACAAATTACGACGAGTGTGGCCGCCGACTGTGAGCGGCTTT 857  
QY 73 TTTGTGTGT 81  
DB 856 TTTCTCTCT 848

RESULT 8  
US-08-951-844-5/c  
Sequence 5, Application US/08951844  
Patent No. 6074860  
GENERAL INFORMATION:  
APPLICANT: Robertson et al.  
TITLE OF INVENTION: Catalases  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
ADDRESS: CECCHI, STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/951,844  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/674,887  
FILING DATE: July 3, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Charles J. Herion  
REGISTRATION NUMBER: 28,019  
REFERENCE/DOCKET NUMBER: 331400-55  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2262 NUCLEOTIDES  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: Oligonucleotide  
US-08-951-844-5

Query Match 20.0%; Score 25.8; DB 3; Length 2262;  
Best Local Similarity 60.9%; Pred. No. 5.9; Mismatches 27; Indels 0; Gaps 0;  
Matches 42; Conservative 0;

QY 13 CGTAGCATTTACTTCGACGATTACGACACCAAGACCCGACGAGATGTCGGGGTCT 72  
DB 916 CATTCGCAATTACCGGACAAATTACGACGAGTGTGGCCGCCGACTGTGAGCGGCTTT 857  
QY 73 TTTGTGTGT 81  
DB 856 TTTCTCTCT 848

RESULT 9  
US-09-412-347-5/c  
Sequence 5, Application US/09412347  
Patent No. 6410290  
GENERAL INFORMATION:  
APPLICANT: Robertson, Dan E.  
APPLICANT: Sanyal, Indrajit  
APPLICANT: Adhikari, Robert S.  
TITLE OF INVENTION: CATALASES  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: LA JOLLA  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/412,347  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/674,887  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Hallie, Ph.D., Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09015/002001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5099  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:

LENGTH: 2262 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 1...2259  
US-09-412-347-5

Query Match 20.0%; Score 25.8; DB 4; Length 2262;  
Best Local Similarity 60.9%; Pred. No. 5.9;  
Matches 42; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 13 GATACGATTACTTGCAGCATTAAGACCCGACGAGATGCGGGTCT 72  
DB 916 CATTCGCATTACCGTACGACATTAACGAGGTGTGCGCCGACGCTGAGGGCTT 857  
QY 73 TTTGCTGT 81  
DB 856 TTTGCTGT 848

## RESULT 10

US-08-461-244-1/C  
Sequence 1, Application US/08461244  
Patent No. 5776729

## GENERAL INFORMATION:

APPLICANT: Soppet, Daniel R.  
APPLICANT: Y1, L1  
APPLICANT: Ruben, Steven M.  
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HGBR32  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:

ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,  
ADDRESSEE: STUART & OLSTEIN  
STREET: 6 Becker Farm Road  
CITY: Roseland  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07068

## COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/461,244  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:  
NAME: Ferraro, Gregory D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-445  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1586 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:

NAME/KEY: CDS  
LOCATION: 431...1495  
US-08-461-244-1

Query Match 19.7%; Score 25.4; DB 1; Length 1586;  
Best Local Similarity 68.6%; Pred. No. 7.1;  
Matches 35; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 35 ACTGACGAAGACCCGACGAGATGTGCGGCTTTTGTGTGTCG 85  
DB 1165 AATGACCAATGACGACCACTGATGCGCTTGTGTGTGTTTG 1115

## RESULT 11

US-09-221-017B-174  
Sequence 174, Application US/09221017B  
Patent No. 6444799

## GENERAL INFORMATION:

APPLICANT: Ross, Bruce C.  
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
NUMBER OF SEQUENCES: 1120  
CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOENSTER  
STREET: 755 PAGE MILL ROAD  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/221,017B  
FILING DATE: 23-DEC-1998  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1182  
FILING DATE: 31-DEC-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1546  
FILING DATE: 30-JAN-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP2911  
FILING DATE: 09-APR-1998  
PRIOR APPLICATION DATA:

ATTORNEY/AGENT INFORMATION:  
NAME: Morroy, Gladys H.  
REGISTRATION NUMBER: 32,430  
REFERENCE/DOCKET NUMBER: 27340-20021.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-813-5600  
TELEFAX: 650-494-0792  
TELEX: 706141

## INFORMATION FOR SEQ ID NO: 174:

SEQUENCE CHARACTERISTICS:  
LENGTH: 3302 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: UNKNOWN  
ORIGINAL SOURCE:  
ORGANISM: PORPHYROMONAS GINGIVALIS  
FEATURE:

NAME/KEY: misc-feature  
LOCATION: 1...3302  
US-09-221-017B-174

Query Match 19.7%; Score 25.4; DB 4; Length 3302;  
Best Local Similarity 54.9%; Pred. No. 9.6;  
Matches 50; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 2 GACCTCTGCTGCTGATTAAGATTAAGACCAAGACCCGACGAGAT 61  
DB 921 GAATTCAGCTCAGCTCAGCTGACGATGCTGATCTCCAGCATTCGCGCTGAT 980

QY 62 GGTGGGGCTTTTGTGTGCTGCTGAC 92  
| | | | | | | | | | | | | | | | | |  
Db 981 TTGCTGGCGCATATGATGATGCTCTGCC 1011

## RESULT 12

US-08-247-901C-1  
; Sequence 1, Application US/08247901C  
; Patent No. 5750384

## GENERAL INFORMATION:

APPLICANT: Jacobs et al  
TITLE OF INVENTION: L5 SHUTTLE PHASMIDS  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amster, Rothstein & Edenstein  
STREET: 90 Park Avenue  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10016

## COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Word Processor (ASCII)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/247,901C  
FILING DATE: May 23, 1994  
CLASSIFICATION: 435

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/057,531  
FILING DATE: April 29, 1993

## ATTORNEY/AGENT INFORMATION:

NAME: Bogosian, Elizabeth A

REGISTRATION NUMBER: 39,911

REFERENCE/DOCKET NUMBER: 96700/273

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 697-5995

TELEFAX: (212) 286-0854 or 286-0082

TELEX: TWX 710-581-4766

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 50341

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: L5 shuttle plasmid sequence

DESCRIPTION: NO

HYPOTHETICAL: NO

ANTI-SENSE:

FRAGMENT TYPE:

ORIGINAL SOURCE:

ORGANISM: L5 mycobacteriophage

STRAIN:

INDIVIDUAL ISOLATE:

DEVELOPMENTAL STAGE:

HAPLOTYPE:

TISSUE TYPE:

CELL TYPE:

CELL LINE:

ORGANELLE:

IMMEDIATE SOURCE:

POSITION IN GENOME:

CHROMOSOME/SEGMENT:

FEATURE:

NAME/KEY:

LOCATION:

IDENTIFICATION METHOD:

OTHER INFORMATION:

PUBLICATION INFORMATION: No. 5750384e

AUTHORS:

TITLE:

## JOURNAL:

VOLUME:

PAGES:

DATE:

DOCUMENT NUMBER:

FILING DATE:

PUBLICATION DATE:

RELEVANT RESIDUES IN SEQ ID NO:

US-08-247-901C-1

Query Match

Best Local Similarity

Matches

36; Conservative

0; Mismatches

18; Indels

0; Gaps

0;

QY 54 ACCGATGTCGGGCTTTTGTGTGCTGACGTGTGCAACCGT 107

| | | | | | | | | | | | | | | | | |

Db 1676 ACCGTGACATCGGCGACATGCTGATGTCGACGTGCGGCGGT 1929

## RESULT 13

US-09-075-904-1  
; Sequence 1, Application US/09075904  
; Patent No. 5994137

## GENERAL INFORMATION:

APPLICANT: Jacobs, et al.  
TITLE OF INVENTION: L5 SHUTTLE PHASMIDS  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amster, Rothstein & Edenstein  
STREET: 90 Park Avenue  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10016

## COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Word Processor (ASCII)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/075,904  
FILING DATE: May 11, 1998  
CLASSIFICATION:

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/247,901

FILING DATE: May 23, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Bogosian, Elizabeth A

REGISTRATION NUMBER: 39,911

REFERENCE/DOCKET NUMBER: 96700/475

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 697-5995

TELEFAX: (212) 286-0854 or 286-0082

TELEX: TWX 710-581-4766

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 50341

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: L5 shuttle plasmid sequence

DESCRIPTION: NO

HYPOTHETICAL: NO

ANTI-SENSE:

FRAGMENT TYPE:

ORIGINAL SOURCE:

ORGANISM: L5 mycobacteriophage

STRAIN:

INDIVIDUAL ISOLATE:

DEVELOPMENTAL STAGE:

HAPLOTYPE:

TISSUE TYPE:

CELL TYPE:

```

: CELL LINE:
: ORGANELL:
: IMMEDIATE SOURCE:
: POSITION IN GENOME:
: CHROMOSOME/SEGMENT:
: FEATURE:
: NAME/KEY:
: LOCATION:
: IDENTIFICATION METHOD:
: OTHER INFORMATION:
: PUBLICATION INFORMATION: No. 5994137e
: AUTHORS:
: TITLE:
: JOURNAL:
: VOLUME:
: PAGES:
: DATE:
: DOCUMENT NUMBER:
: FILING DATE:
: PUBLICATION DATE:
: RELEVANT RESIDUES IN SEQ ID NO:
: US-09-075-904-1

Query Match 19.5%; Score 25.2; DB 2; Length 50341;
Best Local Similarity 66.7%; Pred. Nismatches 18; Indels 0; Gaps 0
Matches 36; Conservative 0; Mismatches 18; Indels 0; Gaps 0

Oy 54 ACCGAGATGCGGGGCTTTTGTGTGTCGCTGTGACGTGTGTCCACCGT 107
Db 1876 ACCGACACCTGCGGACTCGCTGATGTGAGCTGGAGCTATTCGCCGCGT 1929

RESULT 14
US-09-426-436-1
: Sequence 1, Application US/09426436
: Patent No. 6225066
: GENERAL INFORMATION:
: APPLICANT: William R. Jacobs, Jr.
: APPLICANT: Barry R. Bloom
: APPLICANT: Graham F. Hatfull
: TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC
: TITLE OF INVENTION: REPORTER MYCOBACTERIOPHAGES
: NUMBER OF SEQUENCES: 1
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Amster, Rothstein & Epenstein
: STREET: 90 Park Avenue
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10016
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
: COMPUTER: IBM PC Compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: Word Processor (ASCII)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/426,436
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/705,557
: FILING DATE:
: APPLICATION NUMBER: US/08/057,531
: FILING DATE:
: APPLICATION NUMBER: 07/833,431
: FILING DATE: February 7, 1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Pasqualini, Patricia A.
: REGISTRATION NUMBER: 34,894
: REFERENCE/DOCKET NUMBER: 96700/238
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 697-5995
: TELEFAX: (212) 286-0854 or 286-0082

```

```

      TELLEX: TWX 710-581-4766
      INFORMATION FOR SEQ ID NO: 1:
      SEQUENCE CHARACTERISTICS:
          LENGTH: 52297
          TYPE: nucleotide
          STRANDEDNESS: single
          TOPOLOGY: linear
      MOLECULE TYPE:
          DESCRIPTION: phage genome sequence
          HYPOTHETICAL: no
          ANTI-SENSE: no
          FRAGMENT TYPE: not applicable.
          ORIGINAL SOURCE:
              ORGANISM: mycobacteriophage L5
              STRAIN: not applicable
              INDIVIDUAL ISOLATE: L5
              DEVELOPMENTAL STAGE: not applicable
              HAPLOTYPE: not applicable
              TISSUE TYPE: not applicable
              CELL TYPE: not applicable
              CELL LINE: not applicable
              ORGANELLE: not applicable
              IMMEDIATE SOURCE: mycobacteriophage L5 particles
              POSITION IN GENOME: entire genome
          FEATURE:
              NAME/KEY:
                  LOCATION:
                      IDENTIFICATION METHOD:
                      OTHER INFORMATION:
                          PUBLICATION INFORMATION:
                              AUTHORS: Hatfull and Sarkis
                              TITLE: DNA Sequence, Structure and Gene Expression of Mycobacteriophage L5: A Phage System for Mycobacterial Genetics
                              JOURNAL: Molecular Microbiology
                              VOLUME: 7
                              PAGES: 395-405
                              DATE: 1993
      US-09-426-436-1

Query Match           19.5%; Score 25.2; DB 4; Length 52297;
Best Local Similarity 66.7%; Pred.No. 35;
Matches    36; Conservative   0; Mismatches   18; Indels     0; Gaps     0;

QY       54 ACCGAGATCGCggcgcttttTGTgTGnGCNtgACgtGTGCCAACCCT 107
        ||| | | | | | | | | | | | | | | | | | | | | | | | |
Db       1773 ACCGGACACTCGcgcactccGTatGTcGAcTtGGAcGTAATTCCCGGCCGT 1826

RESULT 15
US-08-705-557-1
Sequence 1, Application US/08705557
Patent No. 6300061
GENERAL INFORMATION:
APPLICANT: William R. Jacobs, Jr.
APPLICANT: Barry R. Bloom
APPLICANT: Graham F. Hatfull
TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC
TITLE OF INVENTION: REPORTER MYCOBACTERIOPHAGES
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amster, Rothstein & Edenstein
STREET: 90 Park Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch 1.44 mb storage diskette
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Word Processor (ASCII)

```

;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/705,557  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/057,531  
;; FILING DATE:  
;; APPLICATION NUMBER: 07/833,431  
;; FILING DATE: February 7, 1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Pasqualini, Patricia A.  
;; REGISTRATION NUMBER: 34,894  
;; REFERENCE/DOCKET NUMBER: 96700/238  
;; TELEPHONE: (212) 697-5995  
;; TELEFAX: (212) 286-0854 or 286-0082  
;; TELEX: TWX 710-581-4766  
;; INFORMATION FOR SPO ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 52297  
;; TYPE: nucleotide  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: phage genome sequence  
;; HYPOTHETICAL: no  
;; DESCRIPTION: phage genome sequence  
;; ANTI-SENSE: no  
;; FRAGMENT TYPE: not applicable.  
;; ORIGINAL SOURCE:  
;; ORGANISM: mycobacteriophage L5  
;; STRAIN: not applicable  
;; INDIVIDUAL ISOLATE: L5  
;; DEVELOPMENTAL STAGE: not applicable  
;; HAPLOTYPE: not applicable  
;; TISSUE TYPE: not applicable  
;; CELL TYPE: not applicable  
;; CELL LINE: not applicable  
;; ORGANELLE: not applicable  
;; IMMEDIATE SOURCE: mycobacteriophage L5 particles  
;; POSITION IN GENOME: entire genome  
;; FEATURE:  
;; NAME/KEY:  
;; LOCATION:  
;; IDENTIFICATION METHOD:  
;; OTHER INFORMATION:  
;; PUBLICATION INFORMATION:  
;; AUTHORS: Hatfull and Sarkis  
;; TITLE: DNA Sequence, Structure and Gene  
;; TITLE: Expression of Mycobacteriophage L5:  
;; TITLE: A Phage System for Mycobacterial  
;; TITLE: Genetics  
;; JOURNAL: Molecular Microbiology  
;; VOLUME: 7  
;; PAGES: 395-405  
;; DATE: 1993  
;; US-08-705-557-1

Query Match: 19.5%; Score 25.2; DB 4; Length 52297;  
Best Local Similarity 66.7%; Pred. No. 35;  
Matches 36; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 54 ACCGAGATGTCGGGGCTTTTGTGTGTGTGACGTGTGTCACCGT 107  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1773 ACCGTGACATCGGACATCGCTGATGTCGAGCTGGACGTATTCCCGGCGT 1826

Search completed: December 25, 2002, 12:59:44  
Job time : 89.3364 secs





Db 592 AGTCACGG 600

RESULT 2

US-10-027-805-31  
Sequence 31, Application US/10027805  
Patent No. US20020164725A1

GENERAL INFORMATION:

APPLICANT: ROBERTSON, Daniel E.

MURPHY, Dennis

REID, John

MAFFIA, Anthony

LINK, Steven

SWANSON, Ronald V.

WARREN, Patrick V.

KOSMOTKA, Anna

TITLE OF INVENTION: ESTERASES

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESS: FISH & RICHARDSON P.C.

STREET: 4225 EXECUTIVE SQUARE, STE 1400

CITY: LA JOLLA

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH DISKETTE

COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS

SOFTWARE: WORD PERFECT 6.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/027,805

FILING DATE: 21-Dec-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/602,359

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: HATLE, LISA A.

REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: 09010/010001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-678-5070

TELEFAX: 619-678-5099

INFORMATION FOR SEQ ID NO: 31:

SEQUENCE CHARACTERISTICS:

LENGTH: 936 NUCLEOTIDES

TYPE: NUCLEIC ACID

STRANDEDNESS: SINGLE

TOPOLOGY: LINEAR

MOLECULE TYPE: GENOMIC DNA

SEQUENCE DESCRIPTION: SEQ ID NO: 31:

US-10-027-805-31

Query Match 22.0%; Score 28.4; DB 9; Length 936;

Best Local Similarity 56.4%; Pred. No. 0.74; Indels 0; Gaps 0;

Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Db 718 GACCTTCTGCTAGAGATTACTTGCAGCATCTGACGACCAAGACCCGACCGAGAT 61

Db 718 GACCTTCTGAGAACTTCTGCTGCGCTGATCATTAACCGCAATACGACCGCTGAGAGAT 777

Db 62 GGTGGGGGCTTTTGTGTGCTGCTGACGCG 95

Db 778 GAAGGAGAACTTTTCGGGCGAGATGCTGAGAAAG 811

RESULT 3

US-09-903-410-31  
Sequence 31, Application US/09903410  
Patent No. US20020146799A1

GENERAL INFORMATION:

APPLICANT: DIVERSA CORPORATION

APPLICANT: ROBERTSON, Dan

APPLICANT: MURPHY, Dennis

APPLICANT: REID, John

APPLICANT: MAFFIA, Anthony

APPLICANT: LINK, Steven

APPLICANT: SWANSON, Ronald

APPLICANT: WARREN, Patrick

APPLICANT: KOSMOTKA, Anna

TITLE OF INVENTION: ENZYMES HAVING ESTERASE ACTIVITY AND METHODS OF USE THEREOF

FILE REFERENCE: DIVER1180-2

CURRENT APPLICATION NUMBER: US/09/903,410

CURRENT FILING DATE: 2001-07-10

PRIOR APPLICATION NUMBER: US 09/382,242

PRIOR FILING DATE: 1999-08-24

PRIOR APPLICATION NUMBER: US 08/602,359

PRIOR FILING DATE: 1996-02-16

NUMBER OF SEQ ID NOS: 42

SOFTWARE: PatentIn version 3.0

SEQ ID NO 31

LENGTH: 936

TYPE: DNA

ORGANISM: Archaeoglobus fulgidus

US-09-903-410-31

Query Match 22.0%; Score 28.4; DB 10; Length 936;

Best Local Similarity 56.4%; Pred. No. 0.74; Indels 0; Gaps 0;

Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Db 718 GACCTTCTGCTAGAGATTACTTGCAGCATCTGACGACCAAGACCCGACCGAGAT 61

Db 718 GACCTTCTGAGAACTTCTGCTGCGCTGATCATTAACCGCAATACGACCGCTGAGAGAT 777

Db 62 GGTGGGGGCTTTTGTGTGCTGCTGACGCG 95

Db 778 GAAGGAGAACTTTTCGGGCGAGATGCTGAGAAAG 811

RESULT 4

US-09-938-842A-64  
Sequence 64, Application US/09938842A  
Patent No. US20020160378A1

GENERAL INFORMATION:

APPLICANT: Harper, Jeff

APPLICANT: Kreps, Joel

APPLICANT: Wang, Xun

APPLICANT: Zhu, Tong

TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

FILE REFERENCE: SCRIPI300-3

CURRENT APPLICATION NUMBER: US/09/938,842A

CURRENT FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/227,866

PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: US 60/264,647

PRIOR FILING DATE: 2001-01-16

PRIOR APPLICATION NUMBER: US 60/300,111

PRIOR FILING DATE: 2001-06-22

NUMBER OF SEQ ID NOS: 5379

SEQ ID NO 64

LENGTH: 2451

TYPE: DNA

ORGANISM: Arabidopsis thaliana

US-09-938-842A-64

Query Match 22.0%; Score 28.4; DB 9; Length 2451;

Best Local Similarity 53.6%; Pred. No. 1.1; Indels 51; Gaps 0;

Matches 59; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Db 15 TAGGATTACTTGCAGCATCTGACGACCAAGACCCGACGAGATGCGGGGCTTTT 74

Db 791 TGCGTCTCTGCGAGATTCGCTTTACAAAGCTCTCAACGACTTTTGAGGCTTTG 850

Db 75 TTGTTGTGCTGTGACGTGTTGTCACCGCTATTATTCCGACTAGTT 124

Db 851 TACTGATGTGGCTGCATTCATGAGCTGTTCATGATGAGATAGAT 900

RESULT 5  
US-10-052-586-591/c

Sequence 591, Application US/10052586  
Patent No. US20020127584A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Chen, Jian

APPLICANT: Desnoyers, Luc

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Pan, James

APPLICANT: Smith, Victoria

APPLICANT: Matnabe, Colin K.

APPLICANT: Wood, William I.

APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: P3430R1C1

CURRENT APPLICATION NUMBER: US/10/052,586

PRIOR APPLICATION NUMBER: 60/059263

PRIOR FILING DATE: 1997-09-18

PRIOR APPLICATION NUMBER: 60/059266

PRIOR FILING DATE: 1997-09-18

PRIOR APPLICATION NUMBER: 60/062250

PRIOR FILING DATE: 1997-10-17

PRIOR APPLICATION NUMBER: 60/063120

PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/063121

PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/063486

PRIOR FILING DATE: 1997-10-21

PRIOR APPLICATION NUMBER: 60/063540

PRIOR FILING DATE: 1997-10-28

PRIOR APPLICATION NUMBER: 60/063541

PRIOR FILING DATE: 1997-10-28

PRIOR APPLICATION NUMBER: 60/063544

PRIOR FILING DATE: 1997-10-28

PRIOR APPLICATION NUMBER: 60/063564

PRIOR FILING DATE: 1997-10-28

PRIOR APPLICATION NUMBER: 60/063734

PRIOR FILING DATE: 1997-10-29

PRIOR APPLICATION NUMBER: 60/063870

PRIOR FILING DATE: 1997-10-31

PRIOR APPLICATION NUMBER: 60/064103

PRIOR FILING DATE: 1997-10-31

PRIOR APPLICATION NUMBER: 60/065311

PRIOR FILING DATE: 1997-11-13

PRIOR APPLICATION NUMBER: 60/066120

PRIOR FILING DATE: 1997-11-21

PRIOR APPLICATION NUMBER: 60/066466

PRIOR FILING DATE: 1997-11-24

PRIOR APPLICATION NUMBER: 60/066772

PRIOR FILING DATE: 1997-11-24

PRIOR APPLICATION NUMBER: 60/069335

PRIOR FILING DATE: 1997-12-11

PRIOR APPLICATION NUMBER: 60/069425

PRIOR FILING DATE: 1997-12-12

PRIOR APPLICATION NUMBER: 60/069870

PRIOR FILING DATE: 1997-12-17

PRIOR APPLICATION NUMBER: 60/068017

PRIOR FILING DATE: 1997-12-18

PRIOR APPLICATION NUMBER: 60/077450

PRIOR FILING DATE: 1998-03-10

PRIOR APPLICATION NUMBER: 60/077632

PRIOR FILING DATE: 1998-03-11

PRIOR APPLICATION NUMBER: 60/077649

PRIOR FILING DATE: 1998-03-11

PRIOR APPLICATION NUMBER: 60/078886  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078939  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079664  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079786  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/080107  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080194  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080327  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080333  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/081049  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081070  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081195  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081838  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082568  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082569  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082704  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082797  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/083495  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083496  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083499  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083559  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/084366  
PRIOR FILING DATE: 1998-05-05  
PRIOR APPLICATION NUMBER: 60/084414  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 60/084639  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084640  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084643  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/085573  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085579  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085580  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085582  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085700  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/086023  
PRIOR FILING DATE: 1998-05-18  
PRIOR APPLICATION NUMBER: 60/086392  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/086486  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/087098  
PRIOR FILING DATE: 1998-05-28  
PRIOR APPLICATION NUMBER: 60/087208  
PRIOR FILING DATE: 1998-05-28  
PRIOR APPLICATION NUMBER: 60/087609

```

PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088722
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088740
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088811
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088825
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088863
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089090
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089908
Query Match 21.6%; Score 27.8; DB 12; Length 2037;
Best Local Similarity 57.5%; Pred. No. 1.6;
Matches 50; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
OY 42 ACAAGACCCGACGAGTGGGCTTTTGTGTGCTGTACGTGTGTC 101
DB 1335 AAAAGACCATATCTACTTGTTCAGGCTCATATATCTGTTGCTGAGCTGGGTTTCC 1276
OY 102 AACCGTATTATTCGAGACTAGTTCAGC 128
DB 1275 CTGTTGCTATCTGAGACTATATGATC 1249
```

```

RESULT 6
US-09-785-770A-15/C
Sequence 15, Application US/09785770A
Patent No. US20020103360A1
GENERAL INFORMATION:
APPLICANT: Pan, Yang
TITLE OF INVENTION: A NOVEL PROTEIN RELATED TO MELANOMA-INHIBITING PROTEIN
FILE REFERENCE: 07334-328001
CURRENT FILING DATE: 2001-02-16
PRIOR APPLICATION NUMBER: US 09/387,462
PRIOR FILING DATE: 1999-09-01
PRIOR APPLICATION NUMBER: US 09/145,056
PRIOR FILING DATE: 1998-09-01
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 4.0
SEQ ID NO 15
LENGTH: 5721
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(5721)
US-09-785-770A-15
```

```

Query Match 21.6%; Score 27.8; DB 10; Length 5721;
Best Local Similarity 57.5%; Pred. No. 2.4;
Matches 50; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```

```

OY 42 ACAAGACCCGACGAGTGGGCTTTTGTGTGCTGTACGTGTGTC 101
DB 1335 AAAAGACCATATCTACTTGTTCAGGCTCATATATCTGTTGCTGAGCTGGGTTTCC 1276
OY 102 AACCGTATTATTCGAGACTAGTTCAGC 128
DB 1275 CTGTTGCTATCTGAGACTATATGATC 1249
```

```

RESULT 7
US-09-785-770A-14/C
Sequence 14, Application US/09785770A
Patent No. US20020103360A1
GENERAL INFORMATION:
APPLICANT: Pan, Yang
TITLE OF INVENTION: A NOVEL PROTEIN RELATED TO MELANOMA-INHIBITING PROTEIN
FILE REFERENCE: 07334-328001
CURRENT FILING DATE: 2001-02-16
PRIOR APPLICATION NUMBER: US 09/387,462
PRIOR FILING DATE: 1999-09-01
PRIOR APPLICATION NUMBER: US 09/145,056
PRIOR FILING DATE: 1998-09-01
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 4.0
SEQ ID NO 14
LENGTH: 8121
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: 5'UTR
LOCATION: (1)...(4)
NAME/KEY: CDS
LOCATION: (5)...(5725)
NAME/KEY: 3'UTR
LOCATION: (5726)...(8121)
US-09-785-770A-14
```

```

Query Match 21.6%; Score 27.8; DB 10; Length 8121;
```



```
FILE REFERENCE: 37-21(10297)C
CURRENT APPLICATION NUMBER: US/09/983,965
CURRENT FILING DATE: 2001-10-26
PRIOR APPLICATION NUMBER: US 09/465,231
PRIOR FILING DATE: 1999-12-15
PRIOR APPLICATION NUMBER: US 60/113,678
PRIOR FILING DATE: 1998-12-17
NUMBER OF SEQ ID NOS: 5912
SEQ ID NO 773
LENGTH: 406
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
OTHER INFORMATION: Clone ID: 23-LIB188-006-Q1-E1-F3
US-09-983-965-773

Query Match      21.2% Score 27.4; DB 10; Length 406;
Best Local Similarity 59.7%; Pred. No. 1.2;
Matches 46; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 48 ACCCGACGAGATGTCGGGCTTTTGTGTGTCGTGACGTGTGCAACCGT 107
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 278 ATCCCTACCTCTGGGGGATGTTTCTGTGGGCTGTAACTGTCGCCACTTC 219
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 108 ATTATCCGAGCTAGTT 124
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 218 ATCAATGCTACTGCTGT 202
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 12
US-09-799-777-143
Sequence 143, Application US/09799777
Patent No. US20020091244A1
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
Hillman, Jennifer L.
Corley, Neil C.
Guegler, Karl J.
Baugh, Mariah
Sather, Susan
Shah, Purvi
TITLE OF INVENTION: HUMAN SIGNAL PEPTIDE-CONTAINING PROTEINS
NUMBER OF SEQUENCES: 154
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/799,777
FILING DATE: 06-Mar-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/002,485
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BILLINGS, LUCY J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0459 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 143:
SEQUENCE CHARACTERISTICS:
LENGTH: 1864 base pairs
TYPE: nucleic acid
```

```
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SCORNOT04
CLONE: 2964329
SEQUENCE DESCRIPTION: SEQ ID NO: 143 :
US-09-799-777-143

Query Match      21.2% Score 27.4; DB 10; Length 1864;
Best Local Similarity 54.5%; Pred. No. 2.2;
Matches 55; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 24 CTTGAGCATTACTGACGACAAACCCGACGAGATGTCGGGCTTTTGTGTGG 83
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 719 CTGCTGGATTCCTTGATTAATCCCTCACCTCCCTGGCGGATGTTCTTGGG 778
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 84 TGTGTGACGTGTGTCCAAACCGTATTATTCGGAGCTAGTT 124
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 779 GGCTGTAACTGTGGCCCACTTCAATGCTTAACCTGGT 819
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 13
US-09-974-298-157
Sequence 157, Application US/09974298
Patent No. US20020156263A1
GENERAL INFORMATION:
APPLICANT: Chen, Huel-Mei
TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
FILE REFERENCE: PA-0037 P
CURRENT APPLICATION NUMBER: US/09/974,298
CURRENT FILING DATE: 2001-10-04
PRIOR APPLICATION NUMBER: 60/238,331
PRIOR FILING DATE: 2000-05-10
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PERL Program
SEQ ID NO 157
LENGTH: 2096
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. US20020156263A1 230488.25
US-09-974-298-157

Query Match      21.2% Score 27.4; DB 9; Length 2096;
Best Local Similarity 54.5%; Pred. No. 2.3;
Matches 55; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 24 CTTGAGCATTACTGACGACAAACCCGACGAGATGTCGGGCTTTTGTGTGG 83
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 719 CTGCTGGATTCCTTGATTAATCCCTCACCTCCCTGGCGGATGTTCTTGGG 778
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 84 TGTGTGACGTGTGTCCAAACCGTATTATTCGGAGCTAGTT 124
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 779 GGCTGTAACTGTGGCCCACTTCAATGCTTAACCTGGT 819
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 14
US-09-815-242-6127
Sequence 6127, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: EUTRA.011a
```

```
;; CURRENT APPLICATION NUMBER: US/09/815,242
;; CURRENT FILING DATE: 2001-03-21
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 6127
;; LENGTH: 2628
;; TYPE: DNA
;; ORGANISM: Escherichia coli
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(2628)
US-09-815-242-6127
```

```
Query Match 21.1%; Score 27.2; DB 10; Length 2628;
Best Local Similarity 52.7%; Pred. No. 2.9;
Matches 59; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
```

```
OY 7 TCTGCTGAGCATTAATCTGAGCATTAATCTGAGCAAGACCCGACCGAGATGTCG 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 995 TCACATGATGGCAGTGCACCATGTCACCGCAAGATCATGACCTGAAGCATCATCG 1054
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 67 GGGCTTTTGTGTGCTGTGACGTGTGTCACACCGTATTATTCGGA 118
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1055 CGGGCTTGTGTCACCGCGTGAGGTGACCCGTCGTACTATTTCGA 1106
```

```
RESULT 15
US-09-990-017-4/C
;; Sequence 4, Application US/09990017
;; Patent No. US20020115168A1
;; GENERAL INFORMATION:
;; APPLICANT: Gao, Zeren
;; TITLE OR INVENTION: NOVEL PROTEIN ZLMDA2
;; FILE REFERENCE: 00-67
;; CURRENT APPLICATION NUMBER: US/09/990,017
;; CURRENT FILING DATE: 2000-11-21
;; PRIOR APPLICATION NUMBER: US 60/252,374
;; PRIOR FILING DATE: 2000-11-21
;; NUMBER OF SEQ ID NOS: 6
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 4
;; LENGTH: 792
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: degenerate sequence
;; NAME/KEY: misc.feature
;; LOCATION: 12, 15, 36, 39, 42, 45, 54, 60, 63, 66, 75, 78, 84, 93, 99,
;; LOCATION: 102, 105, 108, 111, 114, 117, 120, 123, 126, 144, 147, 153,
;; LOCATION: 162, 168, 174, 180, 183, 186, 189, 198, 204, 210, 213, 219,
;; LOCATION: 222, 228, 231, 234, 240, 246, 249, 252, 258, 261, 264
;; OTHER INFORMATION: n = A,T,C or G
;; NAME/KEY: misc.feature
;; LOCATION: 273, 279, 282, 294, 300, 303, 306, 309, 321, 327, 345, 372,
;; LOCATION: 378, 384, 393, 396, 399, 402, 405, 408, 423, 426, 438, 444,
;; LOCATION: 447, 450, 468, 474, 483, 486, 498, 504, 510, 513, 516, 525,
;; LOCATION: 528, 531, 534, 537, 543, 549, 552, 564, 585, 591, 594
;; OTHER INFORMATION: n = A,T,C or G
```

```
;; NAME/KEY: misc.feature
;; LOCATION: 597, 618, 630, 648, 651, 654, 657, 660, 663, 678, 696, 699,
;; LOCATION: 702, 705, 708, 711, 714, 717, 720, 723, 729, 738, 750, 756,
;; LOCATION: 765, 771, 777, 780, 783, 789, 792
;; OTHER INFORMATION: n = A,T,C or G
US-09-990-017-4
```

```
Query Match 20.8%; Score 26.8; DB 10; Length 792;
Best Local Similarity 39.7%; Pred. No. 2.6;
Matches 31; Conservative 18; Mismatches 29; Indels 0; Gaps 0;
```

```
OY 4 CCTTCTGCTGAGCATTAATCTGAGCATTAATCTGAGCAAGACCCGACCGAGATG 63
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 643 CTTTTRTCRCNCKRFGDATCATNACRTCATTTATRTCTTTCNCSMDATING 584
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 64 TCGGGCTCTTTTGTTC 81
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 583 TTTGRTTCTTTTCTTTT 566
```

```
Search completed: December 25, 2002, 12:34:36
Job time : 39.9439 secs
```